J. Mol. Biol. (1989) **210**, 347–367

Molecular Dynamics Refinement of a Thermitase–Eglin-c Complex at 1·98 Å Resolution and Comparison of Two Crystal Forms that Differ in Calcium Content

P. Gros[1], Ch. Betzel[2], Z. Dauter[2]
K. S. Wilson[2] and W. G. J. Hol[1]

[1]*Laboratory of Chemical Physics*
University of Groningen
Nijenborgh 16
9747 AG Groningen
The Netherlands

[2]*European Molecular Biology Laboratory, c/o DESY*
Notkestrasse 85
2000 Hamburg 52
West-Germany

(Received 24 April 1989)

The crystal structure of the complex of thermitase with eglin-c in crystal form II, obtained in the presence of 5 mM-$CaCl_2$, has been determined at 1·98 Å resolution. The structure was solved by a molecular replacement method, then molecular dynamics crystallographic refinement was started using the thermitase–eglin-c structure as determined for crystal form I. A ten degrees rigid body misplacement of the core of eglin-c was corrected by the molecular dynamics crystallographic refinement without any need for manual rebuilding on a graphics system. A final crystallographic *R*-factor of 16·5% was obtained for crystal form II.

The comparison of the complexes of thermitase with eglin-c in the two crystal forms shows that the eglin-c cores are differently oriented with respect to the protease. The inhibiting loop of eglin binds in a similar way to thermitase as to subtilisin Carlsberg. A tryptophanyl residue at the S_4 site explains the preference of thermitase for aromatic residues of the substrate at the P_4 site. The difference in the P_1 binding pocket, asparagine in thermitase instead of glycine in subtilisin Carlsberg, does not change the binding of eglin-c. The preference for an arginyl residue at the P_1 site of thermitase can be explained by the hydrogen bonding with Asn170 in thermitase.

Three ion-binding sites of thermitase have been identified. The strong and weak calcium-binding sites resemble the equivalent sites of subtilisin Carlsberg and subtilisin BPN′, though there are important amino acid differences at the calcium-binding sites. The medium-strength calcium-binding site of thermitase is observed in the subtilisin family for the first time. The calcium is bound to residues from the loop 57 to 66. A difference in chelation is observed at this site between the two crystal forms of thermitase, which differ in calcium concentration. Additional electron density is observed near Asp60 in crystal form II, which has more calcium bound than form I. This density is possibly due to a water molecule ligating the calcium ion or the result of Asp60 assuming two significantly different conformations.

1. Introduction

(a) *Thermitase*

Thermitase is an extracellular thermostable serine protease, isolated from *Thermoactinomyces vulgaris* (Frömmel & Höhne, 1981). The primary sequence of this protein, which is 279 amino acid residues long, has been determined by Meloun *et al.* (1985). Sequence comparisons clearly indicate that thermitase is a member of the subtilisin family of serine proteases, with the Asp-His-Ser catalytic triad. X-ray structures have been determined of subtilisin BPN′ at high (McPhalen *et al.*, 1985*a*; Katz & Kossiakoff, 1986; Bott *et al.*, 1988) and at moderate

0022–2836/89/220347–21 $03.00/0

resolution (Wright *et al.*, 1969; Drenth *et al.*, 1972), of subtilisin Carlsberg (Bode *et al.*, 1987; McPhalen *et al.*, 1985*b*; Neidhart & Petsko, 1988), as well as of mutants of these proteins (Bryan *et al.*, 1986; Katz & Kossiakoff, 1986). Also, a high-resolution structure is available for proteinase K (Betzel *et al.*, 1988*a*). Significant differences exist within this family in, for example, enzyme catalysis (Brömme *et al.*, 1986; Kleine, 1982) and thermostability (Voordouw *et al.*, 1976; Frömmel & Höhne, 1981). Thermitase is known to be an extremely thermostable member of the subtilisin family (Frömmel & Höhne, 1981) and has been the subject of a number of investigations in our laboratories (Dauter *et al.*, 1988; Kelders *et al.*, 1989; Gros *et al.*, 1989). A preliminary structure of the native enzyme has been presented by Teplyakov *et al.* (1989). The crystallographic refinement of thermitase complexed with eglin-c in a different crystal form, called crystal form I, has been reported by Gros *et al.* (1989). The initial solution of the structure of crystal form II has been described by Dauter *et al.* (1988).

The degree of sequence similarity between several subtilisins is given in Table 1. Thermitase is included in a subtilisin subgroup that contains a single free cysteinyl residue. This free cysteine is located near the active site, as can be seen from the sequences, and may play a role in enzyme catalysis. Binding a Hg-derivative to this cysteine inactivates the protein (Teplyakov *et al.*, 1986). A large number of amino acid differences and a smaller number of insertions and deletions are apparent when comparing the thermitase sequence with that of subtilisin Carlsberg. The largest difference between thermitase and the other subtilisins occurs at the N terminus. The N terminus of thermitase is extended by seven residues. A totally new arrangement of the N terminus in thermitase is observed in comparison to the other subtilisins, as revealed by the previous crystallographic studies on thermitase (Gros *et al.*, 1989): glutamine 2 of subtilisins Carlsberg and BPN′, which is a ligand of a strongly bound calcium ion, is replaced by Asp5 from the N-terminal extension in thermitase. Proteinase K, in contrast, lacks the N-terminal calcium-binding site (Betzel *et al.*, 1988*b*). This illustrates the divergence in evolution at this site of the subtilisins.

The termostability of subtilisins is strongly influenced by the presence of calcium ions (Voordouw *et al.*, 1976). This phenomenon is observed also for thermitase (Frömmel & Höhne, 1981). However, unlike subtilisin Carlsberg and BPN′, thermitase binds three calcium ions (Frömmel & Höhne, 1981) instead of two. The most weakly bound calcium ion stabilizes the enzyme against thermal denaturation or autoproteolysis (Frömmel & Höhne, 1981). These authors report that the other two calcium ions bind much more strongly and protect against autolysis and thermal inactivation in extreme conditions. Pantoliano *et al.* (1988) have reported three crystal structures of subtilisin BPN′, which differed in $CaCl_2$ content of the crystallization medium. Although the calcium concentration strongly influences the rate of thermal inactivation of this subtilisin, only minor main-chain or side-chain alterations were found. For proteinase K, however, Bajorath *et al.* (1989) report a concerted set of small movements due to binding of calcium in the strong calcium-binding site. This site is common to all subtilisin structures determined so far, and is known as the weak calcium-binding site in subtilisin Carlsberg and BPN′. The function of the calcium-binding sites, the local changes due to amino acid differences amongst the different serine proteases, and the location and structure of the "new" calcium-binding site in thermitase are of considerable interest.

Crystal form II of the complex of thermitase with eglin-c, discussed in this paper, has been grown in the presence of 5 mM-calcium chloride (Dauter *et al.*, 1988). Crystal form I of the same complex (Gros *et al.*, 1989), was obtained from a medium free of calcium chloride. The thermitase–eglin-c models of crystal form I and II will be referred to as TE1 and TE2, respectively. Comparison of the three calcium-binding sites of thermitase in the two different crystal forms should give insight into the role of these binding sites with respect to thermostability and protection against autoproteolysis.

Table 1

Sequence identities (%) between members of the subtilisin family and their cysteine content

	Sub. Carlsberg	Sub. BPN′	Thermitase	Proteinase K
Sub. Carlsberg	100	69	47	37
Sub. BPN′	69	100	47	37
Thermitase	47	47	100	36
Proteinase K	37	37	36	100
No. of Cys	0	0	1	5
No. of disulphide bridges	0	0	0	2

(b) *Eglin-c*

Eglin-c is a 70 amino acid residue protein inhibitor, isolated from the leech *Hirudo medicinalis* (Seemüller *et al.*, 1981). This small protein is a very potent inhibitor of subtilisin, chymotrypsin, and the human proteinases elastase and cathepsin G (Seemüller *et al.*, 1986). These inhibiting characteristics make eglin-c a potential pharmaceutical tool. Although eglin-c lacks cysteinyl residues, it obeys the mechanism of inhibitory action against serine proteases (Seemüller *et al.*, 1986) described by Laskowski & Kato (1980). Stabilization of the scissile bond occurs *via* Arg51 and Arg53 in this cysteine-independent inhibitor, as observed in crystal structures of eglin-c with subtilisin Carlsberg (Bode *et al.*, 1986; McPhalen *et al.*, 1985*a*). None of the crystal structures of eglin-c complexed with a serine protease reported so far (Bode *et al.*, 1986; McPhalen *et al.*, 1985*a*; Gros *et al.*, 1989) shows a reduced electron density at the reactive site peptide

bond. This indicates that proteolytic cleavage of the inhibitor is not a prerequisite for, or consequence of, inhibitory action. As in the previous studies on subtilisins (Bode *et al.*, 1986; McPhalen *et al.*, 1985*a*) and thermitase (Gros *et al.*, 1989), eglin-c was used to prevent autodegradation of the serine protease during crystallization. By studying eglin-c in the various complexes, detailed insight is gained about its interesting inhibitory characteristics.

(c) *Crystallographic molecular dynamics refinement*

The application of the molecular dynamics algorithm (Hermans, 1985; Karplus & McCammon, 1981) in biomacromolecular crystallographic refinement was first described by Brünger *et al.* (1987). Fujinaga *et al.* (1989) have implemented the crystallographic refinement in the simulation package GROMOS (van Gunsteren & Berendsen, 1987) analogous to Brünger *et al.* (1987). The reduction of the required model building on a computer graphics system in the refinement process is substantial when using this new method (Brünger, 1988; Fujinaga *et al.*, 1989; Gros *et al.*, 1989). Two types of strategy have been reported that are suitable for obtaining large "automatic" model corrections by molecular dynamics (MD†) refinement (Brünger, 1988; Fujinaga *et al.*, 1989; Gros *et al.*, 1989). Brünger (1988) proposes a simulated annealing (SA) protocol for using the crystallographic molecular dynamics refinement. In contrast, we performed MD refinement as in our previous studies on the thermitase–eglin-c complex, form I (Gros *et al.*, 1989); that is, during the refinement the upper resolution limit of the X-ray diffraction data was gradually increased, concomitant with increasing X-ray weights, while working at relatively low temperatures (600 and 300 K; Gros *et al.* (1989), in comparison to temperatures over 1000 K used by Brünger (1988)). No additional positional restraint of the water molecules in the crystal structure model is necessary, nor any adjustment of the stereochemical force fields, when using this low-temperature protocol (Fujinaga *et al.*, 1989; Gros *et al.*, 1989). Of major interest in the present work is the discovery that the MD refinement procedure was able to perform a substantial rigid body motion fully automatically.

2. Materials and Methods

(a) *Crystal growth*

Thermitase was prepared as described (Frömmel *et al.*, 1978) and was provided for this study by Dr W. Höhne of the Humboldt University, Berlin. Eglin-c was a kind gift from Dr H. P. Schnebli of Ciba-Geigy, Basel.

The thermitase–eglin-c crystals were obtained as described (Dauter *et al.*, 1988). Briefly, crystals were grown by hanging-drop vapour diffusion in Linbro tissue-culture dishes. The hanging drops contained 5 mg thermitase/ml, 1·7 mg eglin-c/ml, 37·5 mM-sodium acetate, 5 mM-$CaCl_2$ and 5% (w/v) PEG 4000 at pH 5·2, 20°C. The drops were 15 μl in volume. The molar ratio of enzyme to inhibitor was 1 : 1·3. A total of 60 drops were set up. In most dishes there were no crystals, but in each of 2 drops, a single large crystal with elongated tetrahedral morphology, about 0·8 mm in each dimension, grew on the surface of the glass.

(b) *Data collection and reduction*

One of the 2 large crystals was used for characterization of the space group on the precession camera (Dauter *et al.*, 1988). The space group, as for crystal form I, was $P2_12_12_1$, but the cell dimensions were quite different, $a = 49{\cdot}3$ Å, $b = 67{\cdot}3$ Å and $c = 90{\cdot}5$ Å, with V_m of 2·1 Å^3/dalton (1 Å = 0·1 nm).

The 2nd crystal of crystal form II was used to collect 3-dimensional data to 1·8 Å using the EMBL beam-line X31 at the DORIS storage ring, DESY, Hamburg. The data were recorded at room temperature, on CEA-25 reflex X-ray film with 3 films/pack using a modified Arndt–Wonacott oscillation camera. A total of 90° of data were measured, the crystal rotating around the *c*-axis. Exposure times/pack were 10 to 15 min. Many reflections at low resolution were saturated on the films in the 1·8 Å data set. The same crystal was therefore used to record 90° of low-resolution (2·8 Å) data. Exposure times were 1 to 2 min. Owing to lack of crystals, no data were recorded for the blind region.

Films were digitized using an Optronics Photoscan P-1000 microdensitometer in the absorbance range 0 to 3 and a 50 μm raster size. The MOSFLM film processing package was used in evaluating the films. The 1·8 Å data presented considerable problems during integration, owing to the crystal slipping during data collection. This was explained by the rather generous amount of mother liquor left in the capillary to ensure that the single available crystal of form II did not dry out. Eight different orientation matrices were finally required for successive groups of films. The data processing is summarized in Table 2, and the 1·8 Å data had a final merging *R*-factor ($R_{sym} = \Sigma\Sigma|I - I_i|/\Sigma\Sigma I_i$, where I is the mean value of i intensity measurements I_i) of 8·2%. All data, including

Table 2

Summary of the data collection for crystal form II

Parameter	High-resolution set	Low-resolution set	High + Low
Wavelength (Å)	1·15	1·488	—
Crystal-to-film (mm)	56·0	90·4	—
Number of films	72 × 3	40 × 3	112
Oscillation range (°)	1·25	2·25	—
Resolution limit (Å)	1·8	2·8	1·8
Reflns measured	84,568	27,266	94,298†
Reflns unique	24,789	7611	26,664
R_{sym} (%)‡	8·2	4·2	7·2

† This number is less than the sum of the low and high-resolution data, as the data from the high-resolution set were excluded for resolution less than 3·5 Å.

‡ $R_{sym} = \Sigma\Sigma|I - I_i|/\Sigma\Sigma I_i$.

† Abbreviations used: MD, molecular dynamics; SA, simulated annealing; EMX, energy minimization, including X-ray restraints; MDXREF, molecular dynamics, including X-ray restraints; BFAC, individual temperature factor refinement; CE, subtilisin Carlsberg–eglin-c complex; TE1, thermitase–eglin-c complex form I; TE2, thermitase–eglin-e complex form II.

negative intensity observations, were included in the calculation of the R-factor.

Before collection of the 2·8 Å data, the crystal was dried more carefully in the capillary. As a result, no further slippage was detected and a single orientation matrix was used for the whole 90° of data. The R_{sym} for the 2·8 Å data was 4·2%. The final merging of all the data for the 2 limiting resolutions is summarized in Table 2, and gave an R-factor of 7·2% for 26,664 reflections making up 93% of the complete data to 1·8 Å; 88·4% of these had an intensity greater than 3 standard deviations. Data from the high-resolution set were excluded from the final merging if they had a resolution less than 3·5 Å.

The intensities were converted to structure factor amplitudes using the programme TRUNCATE (French & Wilson, 1978). Inspection of the R_{sym} and the fraction of observed to unobserved reflections led to the data to 1·98 Å only being used in the refinement described in this paper.

(c) *Molecular replacement and first refinement cycles*

The phase problem was solved by molecular replacement using the fast rotation function and R-factor translation search programmes from the CCP4 package. We first used the unrefined co-ordinates of subtilisin Novo (Drenth *et al.*, 1972) from the Brookhaven data bank (Bernstein *et al.*, 1977) as a model. With these, we were unable to find a satisfactory solution.

We subsequently used refined co-ordinates of the subtilisin Carlsberg–eglin-c complex kindly provided by Professor M. James, Edmonton (McPhalen *et al.*, 1985*b*). The orientation and position of the complex in the unit cell were determined unambiguously from the 1st calculation with this model, in spite of the relatively low level of homology (47% identity) between the 2 enzymes. For both the rotation and translation searches, data in the resolution range 3·0 to 10·0 Å were used, with Patterson integration range limits of 3 and 25 Å. About 1200 reflections were used in the rotation function. The highest peak, which corresponded to the correct orientation of the molecule in the cell, was 114 on an arbitrary scale. The next highest peak in the map was 75; 931 reflections were used for the translation function. The lowest, and correct, minimum had an R-factor of 37·3%, the next lowest was 42% and the average value in the map was 47·6%. The relevant sections of the rotation and translation functions are shown in Fig. 1 of Dauter *et al.* (1988). These results show the great advantage of using accurate, highly refined co-ordinates in molecular replacement computations.

The model of subtilisin Carlsberg–eglin-c was placed in the cell of crystal form II of thermitase–eglin-c. It was refined using Hendrickson–Konnert restrained least-squares minimization (Hendrickson & Konnert, 1981), with the use of fast Fourier routines to calculate electron densities and gradients (Finzel, 1987). In 20 cycles using data from 5 to 3 Å, the R-factor dropped from 51·5 to 30·2%. FRODO (Jones, 1978) was then used, on an Evans & Sutherland PS330 graphics system, to place the majority of the correct side-chains for thermitase in the model. A further 16 cycles of least-squares minimization with data from 5 to 2·8 Å resolution gave an R-factor of 25·0%, and an additional 30 cycles with 5 to 2·5 Å data gave a final R-factor of 26·2%. The eglin had much less well-defined density than the thermitase at this stage. The reason for this is discussed below. At this point, the restrained refinement of crystal form II was stopped. The refined model of crystal form I, TE1 (Gros *et al.*, 1989), was superimposed on the current model of crystal form II, TE2, using C^{α} co-ordinates only, and the refinement continued with the combined use of X-ray data and molecular dynamics.

(d) *Molecular dynamics refinement procedures*

The crystallographic refinement was performed using the molecular simulation package GROMOS (van Gunsteren & Berendsen, 1987). This suite of programs has been modified by Fujinaga *et al.* (1989) to include crystallographic refinement. Both energy minimization (EMX), including X-ray restraints, and molecular dynamics (MDXREF), also including X-ray restraints, have been used in the refinement. For individual temperature factor refinement (BFAC), the program FFTREF of Agarwal (1978) has been used. Typically, the EMX and MDXREF runs consisted of 100 to 250 steps. One step in MDXREF corresponded to a 2fs time-step. During MDXREF, the system was coupled to a constant heat bath (Berendsen *et al.*, 1984). The interaction parameters of GROMOS for covalent bond, angle, dihedral, improper dihedral, Lennard–Jones and electrostatic interactions for simulation *in vacuo* were used without any alteration (van Gunsteren & Berendsen, 1987). The cut-off radius for calculating electrostatic interactions was set to 8 Å. The weight for the X-ray energy was approximately equal to the observed root-mean-square difference in $|F_{obs}|$ and $|F_{calc}|$ during the refinement and is depicted in Fig. 1(b). The limit for maximal co-ordinate shift of any atom before recalculating F_{calc} values was dependent on the higher resolution limit of the X-ray data used in the refinement. At 3·0 Å resolution, this was set to 0·2 Å and at 1·98 Å to 0·1 Å. Between recalculations, the harmonic approximation, as given by Jack & Levitt (1978), was used to update the X-ray gradient. The calculations for EMX, MDXREF and BFAC were performed on a CONVEX C1-XP of the BIOSON Research Institute.

An overview of the weights and the resolution range used for EMX, MDXREF and BFAC is given in Fig. 1(b). First, the starting model of the complex was subjected to 300 steps of EMX, using diffraction data between 8·0 and 3·0 Å resolution. The individual atomic temperature factors from TE1 were used as starting temperature factors for TE2. Throughout the refinement, all reflections with $|F_{obs}| > \sigma_{|F_{obs}|}$ within a certain resolution shell were used. The weight on the crystallographic energy term was gradually increased during these 300 steps. After these initial 300 steps of EMX, 400 steps of MDXREF were performed at 600 K; first at 3·0 Å resolution and after 200 steps at 2·5 Å resolution (see point A in Fig. 1). Before the temperature factor refinement, the structure was refined at 2·5 Å resolution towards a local minimum by EMX. All individual B factors were reset to 12·0 Å^2 before the temperature factor refinement. The resolution was increased in the following EMX runs to 2·2 Å and later to 1·98 Å, which is the upper resolution limit of the current data set. After co-ordinate refinement at 1·98 Å resolution, the individual temperature factors were refined; this time, the full resolution range from 8·0 to 1·98 Å was used. Up to this point in the refinement, point C in Fig. 1, no manual rebuilding was carried out.

Here, an electron density map at 1·98 Å resolution was calculated. Using this map, the model was refitted and water positions were identified by a procedure to be described below. The temperature factor was reset to 12·0 Å^2 for those parts of the model that were refitted. The refinement was continued using 8·0 to 1·98 Å resolution data for all subsequent steps. After 100 steps of

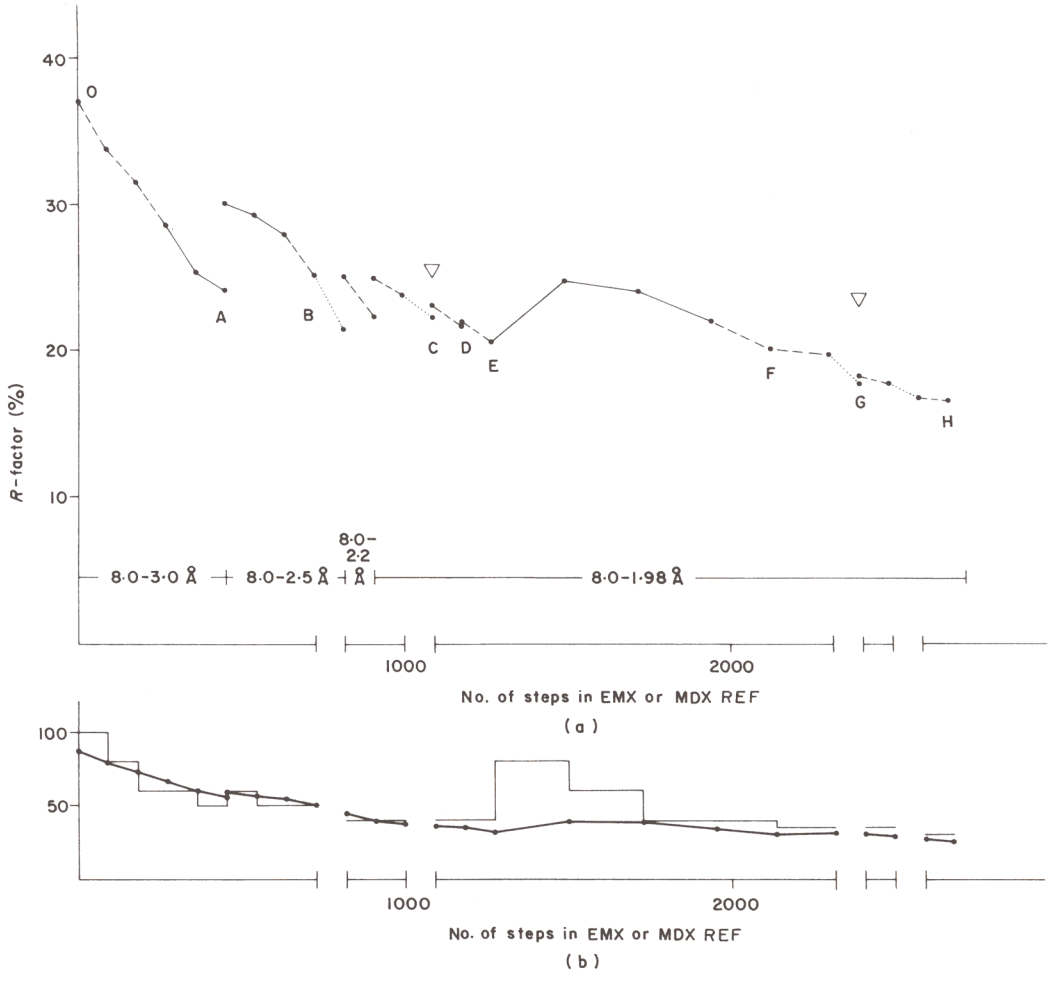

Figure 1. (a) The course of the crystallographic R-factor during the refinement. The R-factor ($R = \Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$) as a function of the number of steps in the refinement is given. Molecular dynamics refinement (MDXREF) is indicated by continuous lines, energy minimization with X-ray restraints (EMX) is given by broken lines and the individual temperature factor refinement is indicated by dots. One step in MDXREF corresponds to a 2fs time-step. Labels indicate points along the refinement referred to in the text. The triangles indicate model building sessions. The resolution range used is shown at the bottom. (b) The root-mean-square difference between $|F_{obs}|$ and $|F_{calc}|$ is given in thick lines. σ_x, used for the calculation of the weight of the X-ray term w_x, is given in thin lines. $w_x = 1/\sigma_{x^2}$ and $E_{x\text{-ray}} = w_x \Sigma(|F_{obs}|-|F_{calc}|)^2$ (Brünger *et al.*, 1987; Fujinaga *et al.*, 1989).

EMX, point D in Fig. 1, the located water molecules were added to the model, all with $B = 14{\cdot}0$ Å^2. Another 100 steps of EMX followed. Here, we chose to reduce the X-ray restraints and use MDXREF again. Remaining potential model errors have a higher chance of being corrected if the restraints are relaxed (Gros *et al.*, 1989). Therefore, MDXREF was performed with a lower weight on the X-ray energy. Gradually, the weight was increased to its previous value. The temperature in MDXREF was this time kept at 300 K, because of the presence of weakly bound water molecules. Before the 2nd and last model building step, the temperature factors were further refined. The last manual corrections were applied to the model and an additional 44 water molecules were positioned; see point G in Fig. 1. This time, the newly found water molecules were given a B of 16·0 Å^2. The resulting model was energy minimized, by 100 steps of EMX. Subsequently, the final individual temperature factors were determined and the co-ordinates further refined by 100 steps of EMX. The ions were identified on the basis of distances to ligands, number of hydrogen bond acceptors as ligands and, for 2 of the 3 sites, homology to other subtilisins. After identification of the ions present in the crystal structure, a few additional steps of EMX were performed with the ions included in the model.

Altogether, 2 sessions of model refitting were carried out on an Evans & Sutherland Picture System 390, using the interactive display program FRODO (Jones, 1978). The density maps used were all calculated with coefficients $(2m|F_{obs}|-D|F_{calc}|)$, $\exp(i\alpha_{calc})$. As shown by Read (1986), the model bias due to co-ordinate errors is suppressed when these coefficients are used. The stereochemical geometry of the model was analysed by calculating the energy/atom throughout the molecules using a program kindly provided by Dr J. Postma. Only the residues that contained energy terms more than 3 to 4σ above the mean energy were inspected and refitted according to the electron density. The water molecules were positioned by locating peaks in the difference electron density map (with coefficients $(m|F_{obs}|-|F_{calc}|),\exp(i\alpha_{calc})$), using the program PEKPIK from the XTAL package (Hall & Stewart, 1987). Only positions with distances between 2·0 and 3·5 Å from any other atom in the model were accepted.

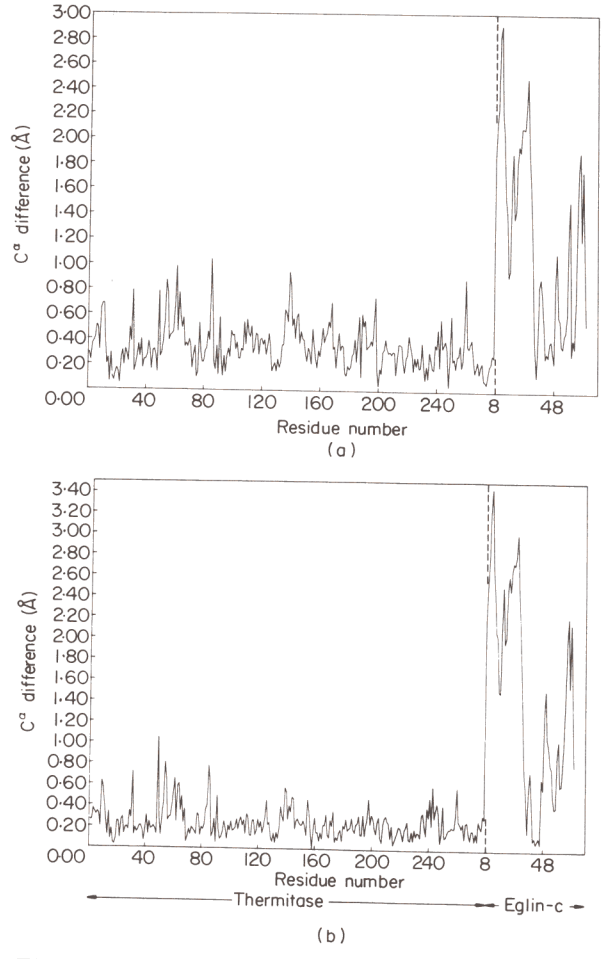

Figure 2. (a) The difference (Å) in C^{α} position between beginning and end of the molecular dynamics refinement (point 0 and point H in Fig. 1) as function of residue number. (b) The difference in C^{α} position between the refined models of TE1 and TE2 as function of residue number. The residue numbering of eglin-c starts with 8, because the 1st 7 residues are invisible in the map and are omitted from the model.

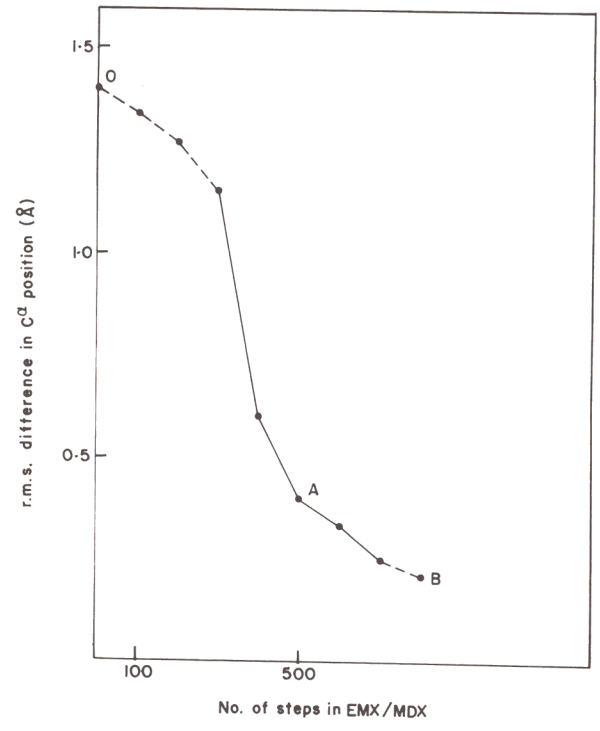

Figure 3. The root-mean-square (r.m.s.) difference in C^{α} positions of eglin-c between intermediate models and the final model shown as a function of the number of steps in EMX or MDXREF needed to obtain the intermediate model. Continuous lines indicate root-mean-square differences during MDXREF runs, and broken lines refer to EMX runs. The labels 0, A and B correspond with the labels in Fig. 1.

3. Results

(a) *Molecular dynamics refinement: accommodation of rigid body motion*

The general decrease of the crystallographic *R*-factor during the refinement is depicted in Figure 1(a). After superposition of TE1 on the preliminary model of the complex in crystal form II, the *R*-factor was 37·0% for data between 8·0 and 3·0 Å resolution; see point 0 in Figure 1(a). A gradual drop in *R*-factor was observed during the EMX, MDXREF and BFAC runs. Before the first model building session, point C in Figure 1, the *R*-factor had dropped to 22·2% for data between 8·0 and 1·98 Å resolution. At point C, the electron density of the model was inspected on a computer graphics system. From the energy analysis, areas with bad geometries were identified. This mainly involved side-chains, such as valine and isoleucine. The worst geometries in the model at point C occurred at Asp62 and Asp188. Asp62 is involved in Ca^{2+} binding, as will be discussed later. Asp188 was known to be less well modelled in the TE1 model, because of its deviating ϕ,φ combination (Gros *et al.*, 1989). In these parts, the model was corrected to fit the electron density as well as possible. The structure was further refined, including 183 water molecules from point D onwards. After point E, the X-ray restraints were relaxed, and thereafter gradually reinforced again to the old value. This process resulted in a decrease in *R*-factor of 0·5% with slightly lower total energies for the stereochemical restraints. At point G in Figure 1, an *R*-factor of 18·0% was obtained for data between 8·0 and 1·98 Å resolution. At this point, 44 new water positions were identified, and the electron density of the model was inspected. The refitting at point G involved: (1) the side-chain of Thr120, from staggered to eclipsed conformation; (2) the side-chains of Val200, Val234, Val335, all from eclipsed to staggered conformation; (3) a slight readjustment of the side-chain of Arg11; and (4) a total refitting of Lys280. By EMX, BFAC and EMX, a final *R*-factor at point H in Figure 1 of 16·5% was obtained, for 19,730 reflections between 8·0 and 1·98 Å resolution, with $|F_{obs}| > \sigma|F_{obs}|$.

Most interestingly, a major model correction had occurred during the refinement, although the

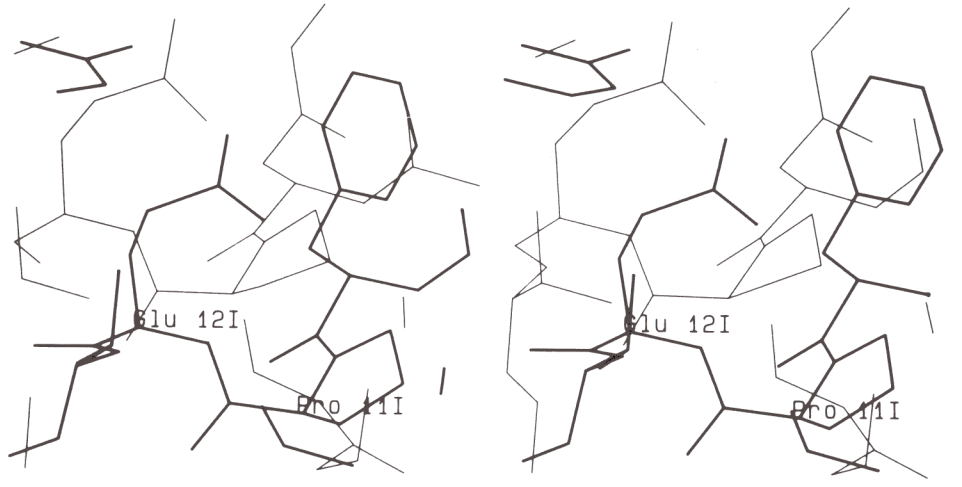

Figure 4. Initial position in thin lines and final position in thick lines of Pro11I and Glu12I of eglin-c giving an impression of the large shifts carried out automatically by the molecular dynamics refinement procedure.

smooth improvement of the *R*-factor, as shown in Figure 1, does not indicate anything striking. Figure 2(a) shows the difference in C^{α} positions for thermitase and eglin-c between point 0 and point H in Figure 1(a). The final differences in C^{α} positions after superposition of the thermitases of the fully refined models is shown in Figure 2(b). The difference in the C^{α} positions of eglin-c is remarkably large. This error in the starting model, at point 0, appeared to correspond to a considerable rigid body motion of the eglin-c core with respect to thermitase.

Figure 3 shows the root-mean-square difference in C^{α} positions for the 63 observed C^{α} atoms of the eglin-c model during the first steps in the refinement with respect to the final C^{α} positions. The root-mean-square deviation between the C^{α} positions of initial and final model was 1·40 Å. The first three runs of EMX brought this down to only 1·15 Å. The effect of increasing the weight (see Fig. 1) on the X-ray energy is reflected in the slopes of these three runs. We refrained from continuing EMX with even stronger weights on the X-ray restraints, because of the increase in the energies for the stereochemical terms during the last 100 steps. Instead, MDXREF was started using the same resolution range, data from 8·0 to 3·0 Å resolution, and the same X-ray weight, $\sigma_x = 60$. The root-mean-square error in C^{α} positions of eglin-c dropped from 1·15 Å to 0·60 Å within the first 100 steps of MDXREF. The second 100 steps of MDXREF with $\sigma_x = 50$ reduced the root-mean-square error to 0·40 Å. Refinement at 2·5 Å resolution brought the error within the limits of accuracy corresponding to the resolution limit; that is, 0·21 Å after MDXREF and EMX with 2·5 Å resolution as upper limit for the data.

The region of eglin-c, where the deviations are among the largest between the starting model (point 0 in Fig. 1) and the final model (point H in Fig. 1), is shown in Figure 4. Clearly, the errors in the initial model are severe. For example, the starting position of Pro11I is closer to the correct position of Glu12I than to its own correct position. The initial position of the C^{α} of Pro11I was 2·2 Å away from the final position of the C^{α} of Glu12I, whereas its own final position was 2·8 Å away from the initial position. For many atoms of eglin-c that are far away from the inhibiting loop, shifts up to 3 Å were performed automatically by the molecular dynamics procedure.

(b) *Accuracy and analysis of the model*

A good indicator of the stereochemical correctness of the main-chain folding of a protein is provided by a Ramachandran plot. Figure 5 displays the angles for thermitase and eglin-c. The allowed regions for non-glycine residues are indicated according to Ramakrishnan & Ramachandran (1965). Large clusters are found at the right-handed α-helix and the β-strand regions. The largest deviation from ideality in φ,ϕ combination for a non-glycine residue is found for Asp38. This is the active site aspartate, which is part of the catalytic triad. Its uncommon φ,ϕ angles are also observed in the other subtilisin structures (Bode *et al.*, 1987; McPhalen *et al.*, 1985*a*,*b*; Betzel *et al.*, 1988*b*), and relate to its catalytic function. There are in total four residues (slightly) outside the strictly allowed regions commonly observed in thermitase and subtilisin Carlsberg; Gln19, Asp38, Asp185 and Thr217. These deviating residues, except Asp38, are all located in short turns. This is probably the reason for the minor deviation in torsion angles from the allowed regions. These deviating φ,ϕ angles of thermitase have been obtained after a substantial exploration of conformational space, which gives the deviations additional relevance. Overall, the Ramachandran plot does not indicate any serious error in the final model of the complex of thermitase and eglin-c.

For the final model of the complex, the energies, as well as the mean energies of the various energy terms, used in the molecular dynamics refinement are given in Table 3. The root-mean-square devia-

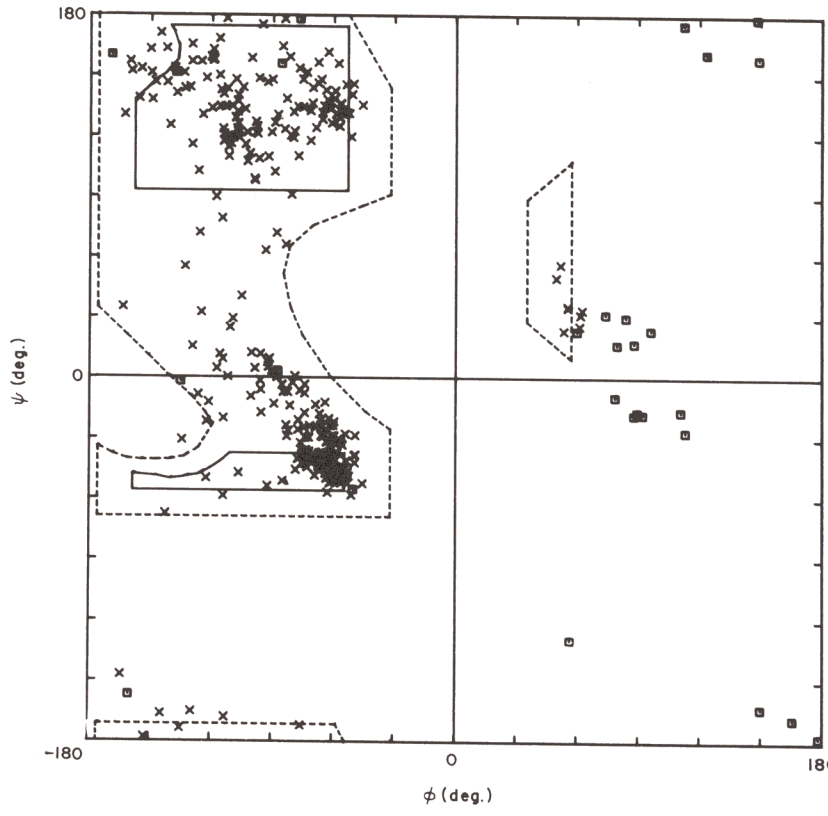

Figure 5. The φ,ϕ plot of the complex of thermitase with eglin-c in crystal form II. The φ,ϕ-angles of non-glycine residues are indicated by a cross and the glycine residues by a square. The continuous lines define the areas of fully allowed conformations with $\tau(N, C^{\alpha}, C) = 110^{\circ}$ for the non-glycine residues. The broken lines show the regions obtained by relaxing the van der Waals' contact constraints as well as by allowing τ to increase to 115° (Ramakrishnan & Ramachandran, 1965).

tions from ideality as calculated by the program TNT (Tronrud *et al.*, 1987) are included. These numbers compare well to what has been obtained for other proteins using more classical least-squares refinement techniques (e.g. see Tronrud *et al.*, 1986). An impression of the accuracy of the model can be obtained from a σ_A plot (Read, 1986), see Figure 6. The overall root-mean-square co-ordinate error for the thermitase–eglin-c complex is 0·13 Å, as estimated from the slope of the σ_A plot.

Table 3

Energies of the final refined model of thermitase–eglin-c complex, form II, and the root-mean-square deviations from ideality

Term	Energy (kJ/mol)
Covalent bond	$0{\cdot}85 \times 10^3$ (0·33)†
Bond angle	$0{\cdot}21 \times 10^4$ (0·58)
Improper dihedral	$0{\cdot}51 \times 10^3$ (0·44)
Dihedral	$0{\cdot}14 \times 10^4$ (0·91)
Lennard–Jones	$-0{\cdot}10 \times 10^5$
Electrostatic	$-0{\cdot}12 \times 10^5$
	Root-mean-square deviation from ideality‡
Covalent bond (Å)	0·019
Bond angle (°)	2·73
Torsion angle (°)	16·77
Trigonal atom non-planarity (Å)	0·026
Planar groups (Å)	0·047
Bad contacts (Å)	0·126

† The average energy, i.e. per bond, bond angle, improper dihedral or dihedral, is given in parentheses.

‡ As calculated by the TNT program (Tronrud *et al.*, 1987). No additional refinement with the TNT program of the model, at point H in Fig. 1, was carried out.

Individual atomic isotropic temperature factors were refined using Agarwal's FFTREF (Agarwal, 1978). An average temperature factor of 11·9 Å^2 was obtained for the whole complex, including 227 water molecules. The average temperature factor of the 2004 non-hydrogen atoms of thermitase is 10·0 Å^2, whereas 12·9 Å^2 was obtained as the average temperature factor for eglin-c, containing 522 non-hydrogen atoms. The 227 water molecules have an average B-factor of 24·6 Å^2. The average main-chain temperature factors are depicted in Figure 7. Very low B-factors in thermitase are observed around the catalytic triad, Asp38, His71 and Ser225. In general, the temperature factors are higher when residues are more exposed to the solvent, as expected. The stabilizing effects of calcium binding are not all that clear from the temperature factors shown in Figure 7. Some residues, like Asp5 and Asp47, have indeed a lower B-factor than their surrounding residues. However, not all ligands from the loop 82 to 89 have a low B-factor. The residues surrounding the calcium-

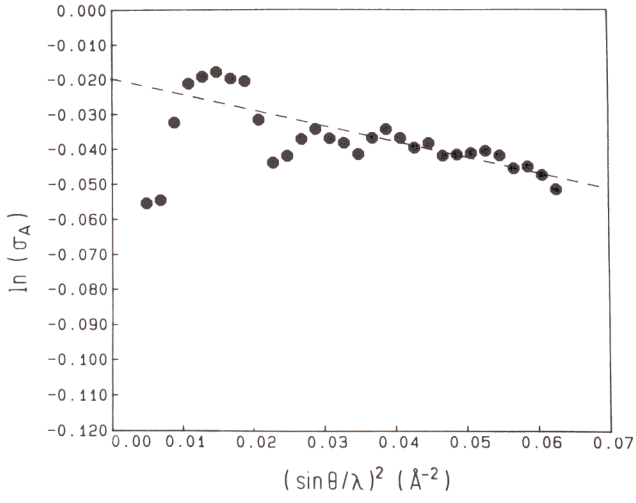

Figure 6. ln (σ_A) plotted against $(\sin\theta/\lambda)^2$. The broken line, fitted by a least-squares procedure, corresponds to a root-mean-square co-ordinate error (Read, 1986) of 0·13 Å. The 1st 2 points were not taken into account in the least-squares fitting.

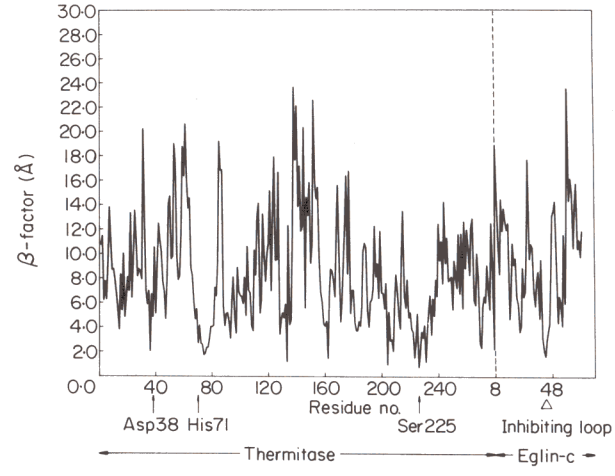

Figure 7. The main-chain temperature factors (A^2), averaged over N, C$^\alpha$, C and O, are given as a function of residue number for thermitase and eglin-c. The catalytic triad residues of thermitase are indicated in the plot, as well as the residues of eglin-c directly involved in binding to thermitase. The residue numbering of eglin-c starts with 8, because the 1st 7 residues are invisible in the map and are omitted from the model.

binding sites will be discussed in more detail in section (d), below.

The lowest *B*-factors for eglin-c are observed in the inhibiting loop, residues 41I to 46I. A second region with low *B*-factors in eglin-c is the stretch 51I to 57I. This stretch contains the inhibiting loop stabilizing residues Arg51I and Arg53I. On average, eglin-c shows a slightly larger flexibility than thermitase, as reflected in the average temperature factors, 12·9 Å^2 and 10·0 Å^2, respectively. This difference in temperature factors between protease and inhibitor is not as large as in the case of subtilisin Carlsberg–eglin-c (Bode *et al.*, 1987), where values of 10·0 Å^2 for subtilisin Carlsberg and 16·5 Å^2 for eglin-c were reported. Although the temperature factors have been refined without restraints, they can clearly be correlated with specific aspects of the atomic model of the complex.

There are two *cis* peptide bonds in the structure of thermitase; both involve prolyl residues. In one case, the peptide precedes a prolyl residue, which is rather common. In the other case, the peptide follows a prolyl residue, which is very rarely observed. Figure 8 shows the two *cis* peptide bonds. The *cis* peptide bond between Pro214 and Thr215 has been observed for subtilisin Carlsberg between the equivalent residues Pro210 and Thr211 (McPhalen *et al.*, 1986). In the two proteins, the *cis* peptide bond facilitates a reverse turn at the surface of the protein. The other *cis* peptide bond found in thermitase is in common with subtilisin Carlsberg and subtilisin BPN′ (Bode *et al.*, 1987; McPhalen & James, 1988) and precedes a prolyl residue. This *cis* proline, Tyr171–Pro172, is shown in Figure 8(b) with its electron density. This *cis* peptide bond can be seen to be required for a short reverse turn in the main chain. This time, the turn is relatively more buried inside the protein. For both *cis* peptide bonds, the electron density is unambiguous.

While the density is well-defined at most places, in a few instances this is not the case. This involves Asp60, Ser138, Val199 and Pro229. Asp60 will be discussed in section (d), below. Ser138 appears to be somewhat disordered in crystal form II, though in crystal form I it has well-defined electron density. Pro229 lacks electron density for its C$^\gamma$ atom. In crystal form I, the full electron density for a prolyl residue is visible. Both crystal forms show electron density that could accommodate a tryptophanyl residue at the position of Val199. The published amino acid sequence might have to be corrected for residue 199 of thermitase.

There are several novel features in the model of thermitase in comparison to the other subtilisins. The largest differences in main-chain fold can be found at insertion and deletion sites. Two of these new features, the N and C-terminal extensions, are highlighted here, see Figure 9. The N-terminal extension, residues Tyr1 to Tyr7, is clearly visible in the electron density map, as shown for residues 3 to 8 in Figure 9(a). Residues Tyr1 to Gln12 are totally rearranged in thermitase with respect to the other subtilisins. A short helical turn consisting of only four residues is formed by Tyr7 to Ser10, following the definition of Kabsch & Sanders (1983). The N-terminal calcium ligand, Gln2 in subtilisin Carlsberg and BPN′, is replaced by Asp5 from the N-terminal extension in thermitase. The terminal α-amino group of thermitase is nicely stabilized by a hydrogen bond to O$^\delta$ of Asp25 (d = 2·82 Å).

The C-terminal extension, Tyr279, is shown in Figure 9(b). Again, the electron density shows the additional residue unambiguously. The C-terminal Tyr279 is stabilized by a hydrogen bond between its hydroxyl oxygen and the side-chain nitrogen of Asn254.

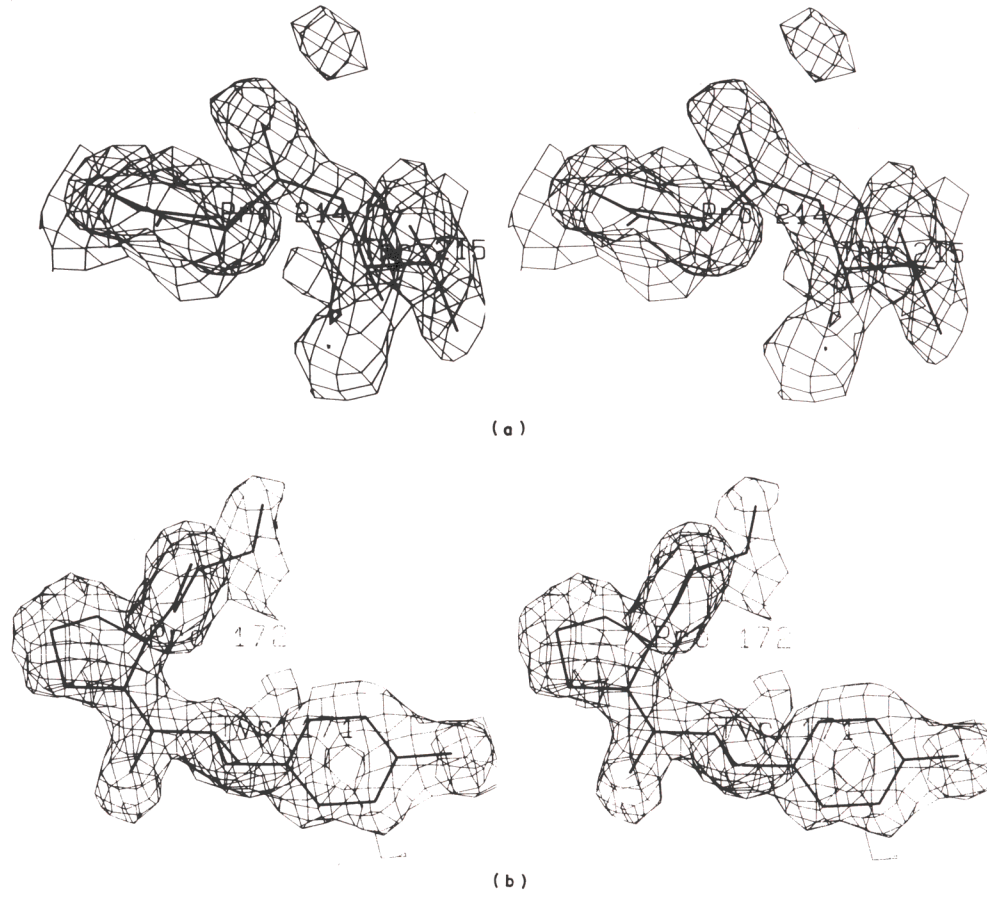

Figure 8. The *cis* peptide bonds in thermitase. (a) The *cis* Thr bond, Pro214–Thr215, is shown in its final $2mF_o$-DF_c,α_c, electron density. (b) The *cis* Pro bond, Tyr171–Tyr172, and the electron density for the 2 residues is shown. The electron densities were contoured at 1·1 σ.

Though there are several novel features in the model, the thermitase core resembles subtilisin Carlsberg to a large extent. The root-mean-square difference in C^{α} co-ordinates between thermitase and subtilisin Carlsberg is 2·02 Å for 267 residues and 1·99 Å for 267 residues between thermitase and subtilisin BPN′. The root-mean-square differences drop to 0·31 Å and 0·50 Å if only the best 148 or 203 superimposable C^{α} atoms are taken into account for subtilisin Carlsberg and BPN′, respectively. These numbers correspond well to what Chothia & Lesk (1986) predict from the degree of amino acid similarity. The structure of the complex thermitase–eglin-c allows a detailed comparison with the other subtilisins for evaluation of the insertions, deletions and many substitutions.

(c) *The binding of eglin-c*

By far the largest difference between the thermitase–eglin-c complexes of TE1 and TE2 is the orientation of the inhibitor core with respect to the protease. Figure 10 shows the C^{α} tracings of TE1 and TE2 after superposition of the thermitase

Table 4

The main-chain torsion angles, φ and ψ, of the inhibiting loop of eglin-c in complex with thermitase, TE1 and TE2, and subtilisin Carlsberg, CE

	Residue	P42I	V43I	T44I	L45I	D46I	L47I	R48I
Complex	P-site	P4	P3	P2	P1	P'_1	P'_2	P'_3
TE1		−68/152	−129/152	−65/133	−99/51	−103/166	−108/129	−137/121
TE2		−71/136	−138/153	−64/143	−105/40	−85/164	−112/110	−110/119
CE		−72/138	−141/165	−58/142	−116/47	−98/169	−118/106	−117/109

ω_{TE1} (L45I−D46I) = 173; ω_{TE2} (L45I−D46I) = 167; ω_{CE} (L45I−D46I) = 161.

The ω angles of the scissile peptide plane are also listed for the 3 complexes. Data concerning subtilisin Carlsberg–eglin-c are taken from Bode *et al.* (1986). All angles are in degrees.

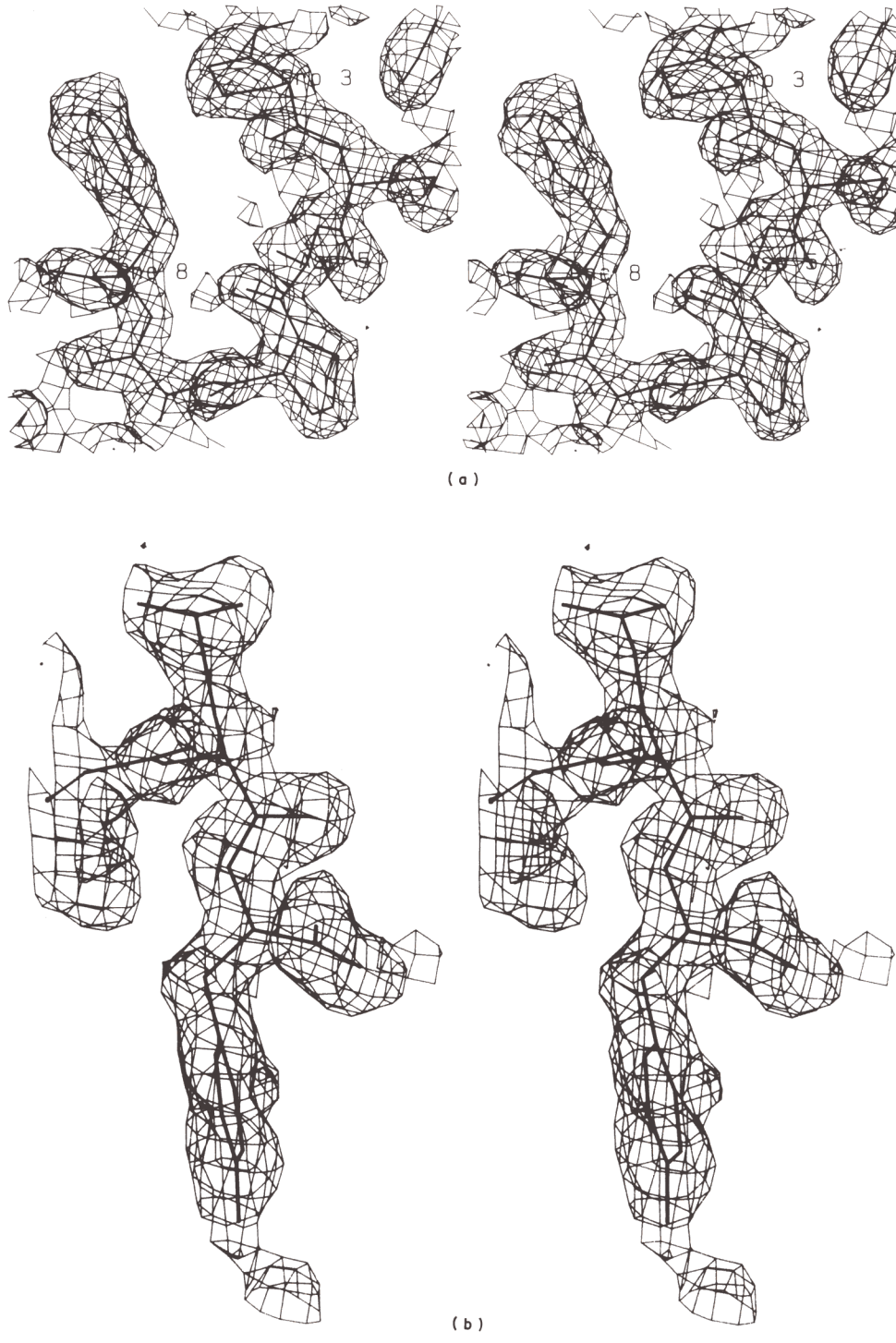

Figure 9. (a) Pro3 to Phe8, from the N-terminal extension of thermitase, shown in the $2mF_o$-DF_c,α_c, electron density. (b) The C-terminal residues Gln278 and Tyr279 of thermitase shown with the electron density. The electron densities are contoured at 1·1 σ.

molecules only. The inhibiting loop of eglin-c, residue Ser41I to Leu47I, is in equivalent positions in both structures. The root-mean-square difference between equivalent C^{α} atoms in this loop is 0·19 Å. The cores of eglin-c, consisting of Lys8I to Gly40I and Arg48I to Gly70I, differ by a rigid body rotation of 10·4°. The root-mean-square difference between the 56 C^{α} atoms of the two cores is 0·42 Å after applying the 10·4° rotation. This shows that the cores by themselves, as well as the inhibiting loops by themselves, are virtually identical and that basically the motion is that of a rigid body with hinges near Ser41I and Leu47I. The structure of eglin-c complexed with subtilisin Carlsberg (Bode *et al.*, 1986; McPhalen *et al.*, 1985*a*) shows, yet again, a slightly different orientation of the eglin-c core. It is

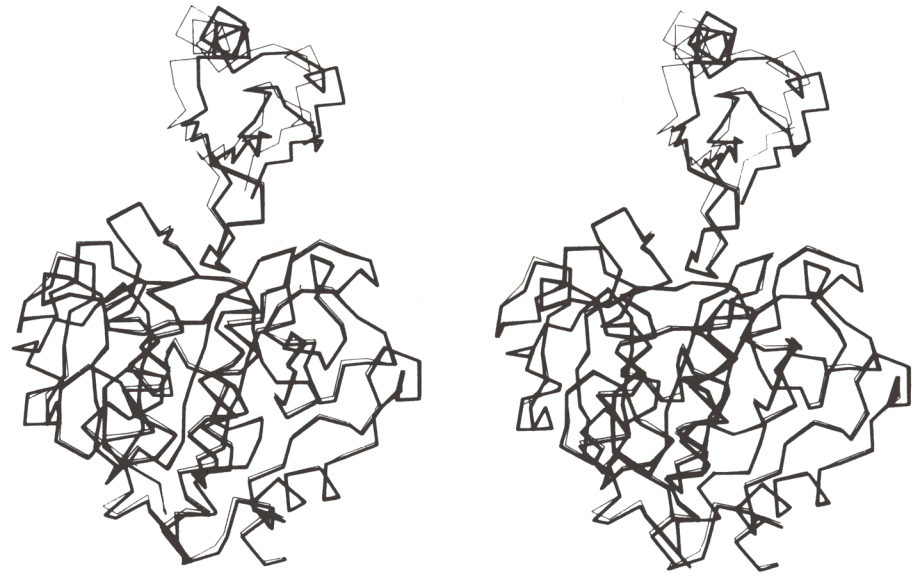

Figure 10. Comparison of the 2 thermitase–eglin-c complexes. The C^{α} tracing of the complex thermitase–eglin-c as observed in crystal form II (TE2) shown in thick lines. The complex in crystal form I (TE1) shown in thin lines. The structure of TE1 was determined by Gros *et al.* (1989) and refined to an *R*-factor of 17·9%. Using the TE2 model, the TE1 model could be improved only slightly. The resulting *R*-factor was 17·5%.

likely that eglin-c is, to some extent, flexible in solution, giving rise to different orientations of its core relative to the protease.

The main-chain torsion angles, φ and ϕ, of the inhibiting loop of eglin-c in TE1, TE2 and in the complex of eglin-c with subtilisin Carlsberg (Bode *et al.*, 1986) are given in Table 4. (For convenience, the complex of subtilisin Carlsberg with eglin-c will sometimes be referred to as CE in this discussion). Clearly, the inhibiting loop exhibits equivalent conformations in the three complexes. The largest distortion from planarity for the scissile bond, Leu45I–Asp46I, is found in the complex with subtilisin Carlsberg. In the complexes with thermitase, a smaller deviation from planarity is observed. The distance between the O^{γ} of Ser225 and the carbonyl carbon of the scissile bond is 2·91 Å and 2·77 Å in TE1 and TE2, respectively, whereas Bode *et al.* (1986) observe a distance of 2·81 Å in the complex of subtilisin Carlsberg with eglin-c. This indicates that proteolysis is stopped at the same point along the reaction pathway in all three complexes.

Table 5

Hydrogen bonding between protease and inhibitor in the complexes of thermitase with eglin-c (TE1 and TE2), and in subtilisin Carlsberg with eglin-c (CE)

Elgin-c		Complex					
		TE1		TE2		CE	
Residue	Atom	Atom	Residue	Atom	Residue	Atom	Residue
Gly40I	O	NE1	Trp112 (2·78; 43·6)	NE1	Trp112 (3·11; 51·5)	OH	Tyr104 (2·57; 26·2)
Pro42I	O			OG	Ser106 (3·46; 9·4)		
		N	Gly110 (3·36; 52·3)	N	Gly110 (3·40; 31·8)	N	Gly102 (3·34; 35·5)
Val43I	N	O	Gly135 (3·22; 10·7)	O	Gly135 (2·88; 24·9)	O	Gly127 (2·97; 7·7)
	O	N	Gly135 (3·00; 31·4)	N	Gly135 (3·02; 21·4)	N	Gly127 (2·94; 21·9)
Thr44I	N	O	Gly108 (3·02; 30·9)	O	Gly108 (2·90; 31·2)	O	Gly100 (2·91; 24·7)
Leu45I	N	O	Ser133 (3·32; 16·4)	O	Ser133 (3·34; 23·3)		
						OG	Ser221 (2·91; 54·8)
	O	ND2	Asn163 (2·76; 5·6)	ND2	Asn163 (2·89; 11·2)	ND2	Asn155 (2·72; 32·5)
		N	Ser225 (3·20; 16·6)	N	Ser225 (3·00; 26·0)	N	Ser221 (2·99; 20·7)
Leu47I	N	O	Ser222 (2·76; 29·4)	O	Ser222 (2·82; 31·7)	O	Asn218 (2·90; 48·7)
Arg48I	NH1	OH	Tyr213 (3·15; 58·5)				
						O	Asn62 (3·00; 19·5)
	NH2					O	Asn62 (3·31; 37·7)
Arg53I	NH2	O	Ser107 (3·47; 15·5)	O	Ser107 (3·36; 21·3)		

The distance (Å) and angular deviation from 180° (°) between hydrogen donor and acceptor are given in parentheses. Only hydrogen bonds with a distance less than 3·5 Å and a deviation less than 60° from linearity are considered. These data were calculated using the program HBONDS within the molecular analysis and graphics program package WHATIF (G. Vriend, unpublished results).

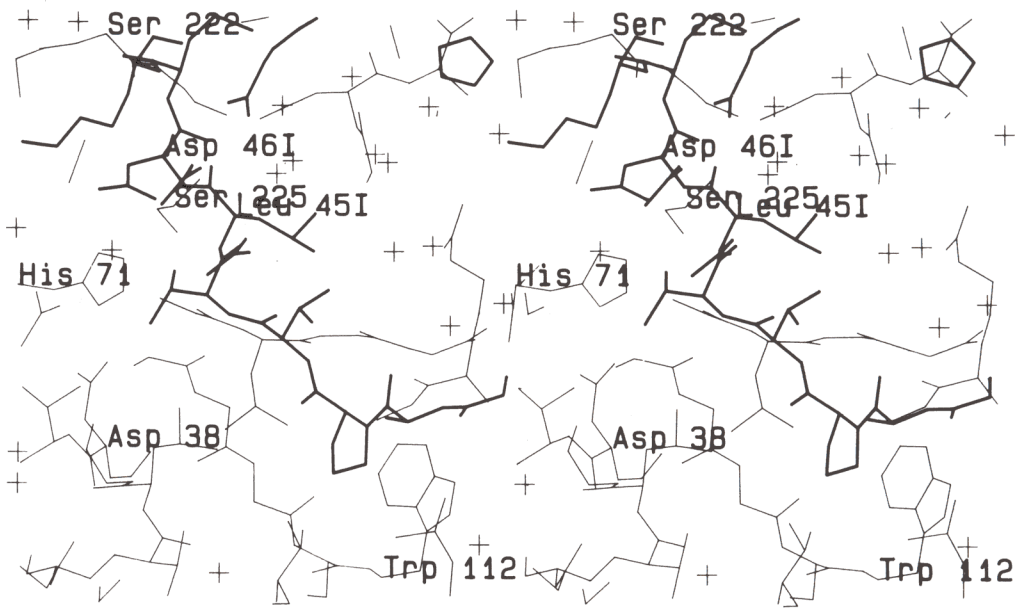

Figure 11. The inhibiting loop of eglin-c bound to thermitase in crystal form II. Residues of eglin-c are shown in thick lines. The active site region and the binding sites of thermitase are shown in thin lines. Labelled residues are discussed in the text. Water molecules are indicated by crosses.

Table 5 shows the hydrogen bonds identified in the current models between the inhibitor, eglin-c, and the two thermitases and subtilisin Carlsberg. The hydrogen-bonding network for Pro42I to Leu47I is very similar for the three cases. Also, the water-mediated contact between Thr44I, Asp46I and Arg51I of eglin-c and Asn62 in subtilisin Carlsberg (Bode *et al.*, 1986) is observed in the two thermitase–eglin-c complexes. The following alterations in the hydrogen pattern can be noted (Table 5).

(1) A slight difference in hydrogen bonding seems to occur for the main-chain nitrogen of Leu45I, when comparing subtilisin Carlsberg with the thermitase complexes. However, the co-ordinate differences between the three complexes are small for the nitrogen of Leu45I and its potential hydrogen acceptors.

(2) Tyr104 in subtilisin Carlsberg is replaced in thermitase by Trp112. The hydrogen bond of the side-chain hydroxyl of Tyr104 with the carbonyl oxygen of Gly40I in the complex of subtilisin Carlsberg with eglin-c is neatly replaced by a hydrogen bond between the side-chain NH group of Trp112 with the same carbonyl oxygen of the inhibitor, in the complex of thermitase with eglin-c.

(3) A rearrangement of hydrogen bonds occurs in the neighbourhood of Arg48I. The carbonyl oxygen of Asn69 in thermitase does not directly hydrogen bond to the inhibitor, unlike the equivalent Asn62 in subtilisin Carlsberg. Instead, it makes a relatively poor hydrogen bond with the hydroxyl of Tyr213. The hydroxyl of Tyr213 has, in turn, a possible hydrogen bond to the $N^{\eta 2}$ of Arg48I.

(4) In TE2, a water molecule mediates between $N^{\eta 2}$ of Arg48I and $O^{\delta 1}$ of Asn69 and the main-chain nitrogen of His71. This water molecule is absent in both TE1 and CE. The insertion of a water molecule in TE2 at the interface of eglin-c with thermitase can be correlated with the rigid body displacement of eglin-c in TE2 compared to eglin-c in TE1 and CE.

The contacts of eglin-c with thermitase and subtilisin Carlsberg are given in Table 6. Of 29 residues of the proteases that are in close contact with eglin-c, 22 are identical. Six of the amino acid substitutions do not influence the P_3 to P'_3 binding pocket directly. The seventh, Ser222 in thermitase (equivalent to Asn218 in subtilisin Carlsberg), is near the P'_1 and P'_2 binding pockets. Nonetheless, this amino acid difference between thermitase and subtilisin Carlsberg does not directly affect the binding, because the side-chains are pointing away from the binding site. Serine, instead of asparagine, at this position is important for the thermostability, as shown by Bryan *et al.* (1986).

There are significant differences in binding of eglin-c in the three complexes. This includes the Tyr104 to Trp112 difference at the S_4 site between subtilisin Carlsberg and thermitase. In the S_4 site, many more contacts of Pro42I with thermitase are observed than with subtilisin Carlsberg. As shown in Figure 11, Trp112 in thermitase stacks nicely with Pro42I, giving rise to many hydrophobic contacts between protease and inhibitor at the S_4 site.

At the S'_3 site, several van der Waals' contacts appear to be lost in the thermitase–eglin-c complexes. As pointed out earlier, the hydrogen bonds are also different at this site, but no amino acid differences are directly involved. The carbonyl oxygen of Asn69 in thermitase is clearly further away from the inhibitor than the equivalent oxygen

Table 6

Number of contacts between eglin-c and thermitase in crystal form I (TE1) and crystal form II (TE2), and between eglin-c and subtilisin Carlsberg (CE)

		P-site						P_5	P_4	P_3	P_2	P_1	P'_1	P'_2	P'_3	P'_4						
		Eglin-c	Tyr		Leu			Gly	Ser	Pro	Val	Thr	Leu	Asp	Leu	Arg	Tyr		Arg		Arg		
			35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	Totals in	
Thermitase	Carlsberg																					TE1	TE2	CE
36 Thr	33 Thr											1,1,1										1	1	1
69 Asn	62 Asn															2,0,8						2	0	8
71 His	64 His											10,10,9	1,1,1	2,2,3		0,0,2						13	13	15
104 Leu	96 Leu											1,1,1										1	1	1
107 Ser	99 Ser		3,1,4																		1,1,1	4	2	5
108 Gly	100 Gly										2,2,2	2,4,3									3,3,3	7	9	8
109 Ser	101 Ser				2,2,0					1,1,2	0,0,1											4	3	3
110 Gly	102 Gly									4,5,4												4	5	4
111 Thr	103 Ser									3,3,0												3	3	0
112 Trp	104 Tyr							4,1,2		15,11,3												19	12	5
115 Val	107 Ile									2,2,2												2	2	2
133 Ser	125 Ser											0,0,1	1,1,0									1	1	1
134 Leu	126 Leu									0,1,0	4,4,5		2,2,3									6	7	8
135 Gly	127 Gly									0,3,2	3,4,3		3,3,2									6	10	7
136 Gly	128 Gly							0,1,0	0,1,0	0,0,1												0	2	1
137 Thr	129 Ala							0,0,5														0	0	5
138 Val	130 Ser							3,2,3														3	2	3
160 Ala	152 Ala												2,1,1									2	1	1
162 Ala	154 Gly												1,2,2									1	2	2
163 Asn	155 Asn												3,4,5	0,1,3								3	5	8
193 Phe	189 Phe														5,3,5							5	3	5
208 Trp	204 Gly																0,6,0					0	6	0
213 Tyr	209 Tyr															3,3,2						3	3	2
221 Leu	217 Leu															0,4,1						0	4	1
222 Ser	218 Asn													2,2,3	3,3,1							5	5	4
223 Gly	219 Gly												2,1,2		1,0,0							5	1	2
224 Thr	220 Thr												1,2,1									1	2	1
225 Ser	221 Ser												9,12,10	2,2,2								11	14	12
226 Met	222 Met													0,0,1								0	0	1
	Totals in																							
	TE1		3		2			7	0	25	9	14	25	6	9	5	0				4			
	TE2		1		2			4	1	26	10	16	29	7	6	7	6				4			
	CE		4		0			10	0	14	11	15	27	12	6	13	0				4			

Two atoms are considered to be in contact when the distance between them is less than the sum of their van der Waals' radii plus 0·5 Å. If contacts occur, each position of this "contact matrix" contains 3 numbers: the 1st being the number of contacts between an inhibitor and a protease residue in TE1, the 2nd being the number in TE2, and 3rd in CE.

of Asn62 in subtilisin Carlsberg. The C^{α} atoms of the asparaginyl residues of thermitase and subtilisin Carlsberg are 0·68 Å apart after superposition. The preceding glycyl residues superimpose even worse, their C^{α} atoms deviate 2·3 Å after superposition. These differences between the proteases are the consequence of the new calcium-binding site in thermitase around residues 57 to 66.

Another interesting amino acid difference, concerning substrate specificity, between thermitase and the other subtilisins is located at the bottom of the S_1 site. In subtilisins Carlsberg and BPN′, Gly166 is found at this position. In the complex of eglin-c with subtilisin Carlsberg, the P_1 pocket is filled with the side-chain of Leu45I and two water molecules. In thermitase, Gly166 is replaced by Asn170. No water molecule can be found in the complexes of thermitase and eglin-c in the P_1 pocket, but also no direct interaction occurs between the side-chains of Leu45I and Asn170. It is likely that the preference for an arginyl residue in the S_1 site (Brömme & Kleine, 1984) may be explained by specific hydrogen bonds of the guanidyl group of arginine with the side-chain of asparagine at position 170 in thermitase.

(d) *The three ion-binding sites*

Biochemical data show that thermitase can bind three calcium ions (Frömmel & Höhne, 1981). All three ion-binding sites have been identified in the electron density map. The residues involved in chelating these ions are listed in Table 7. Two sites, Ca1 and Ca3, are in common with subtilisin Carlsberg and BPN′. In spite of large sequence differences, especially at the N terminus, the chelating ligands are conserved to a considerable extent in these sites. Site Ca2 in thermitase is a newly observed calcium-binding site for the subtilisins. Table 8 gives the distances between ligand and cation for the three calcium-binding sites of both crystal forms of thermitase–eglin-c. Figure 12 shows the three ion-binding sites and their surroundings in crystal form II of the thermitase–eglin-c complex.

The strong calcium-binding site of thermitase, Ca1, has an octahedral arrangement of ligands (Fig. 12(a)). The ligands are provided by three different parts of the protein (Table 7): (1) Asp5 from the N-terminal extension; (2) Asp47; and (3) residues 82, 85, 87 and 89 from the "Ca1-embracing loop". In addition to links *via* the calcium ion, the N-terminal extension and the Ca-1 embracing loop are held together by hydrogen bonds between Asn4 and Asp5 of the extension, and Asn84 of the embracing loop. These interactions between different parts of the protein may contribute to its thermostability.

When comparing crystal forms I and II, no large differences are apparent in the Ca1 site (Table 8). However, Thr87 appears to be 0·18 Å further away from the calcium ion in TE2 than in TE1. This does not correspond with the calcium concentrations in the crystallization medium, which is higher in TE2. The crystallographic *B*-factors of the ions are in both forms equal to the lower limit, which was set to 0·5 $Å^2$ in the refinement. This indicates that the Ca1 site is most likely fully occupied in both crystal forms.

Comparison of the Ca1 site in thermitase with the equivalent, "strong", calcium-binding site in subtilisin Carlsberg and BPN′ (Table 7) shows two differences. First, the Ca1-embracing loop is one residue longer in thermitase. Asn84 in thermitase loops out in comparison to subtilisin Carlsberg, thus

Table 7

Alignment of the thermitase sequence with subtilisin Carlsberg in the regions involved in chelating calcium

A. *Calcium-binding site 1*

	1	2	3	4	5	6	7	8	9	10	11	12	...	47	...	82	83	84	85	86	87	88	89	...
Therm.	Y	T	P	N	D	P	Y	F	S	S	R	Q		D		V	T	N	N	S	T	G	I	
Carlsberg				A	Q					T	V	P		D		L		D	N	T	T	G	V	
				1	2					3	4	5	...	41	...	74		75	76	77	78	79	80	...
					*									*		*			*		*		*	
					S									S		M			S		M		M	

B. *Calcium-binding site 2*†

	...	57	58	59	60	61	62	63	64	65	66	...
Therm.		D	F	V	D	N	D	S	T	P	Q	
Carlsberg		S	F	V	A	G	E	A	Y	N	T	
	...	49	50	51	52	53	54	55	56	57	58	...
		*			*		*		*		*	
		S			S		S		M		S	

C. *Calcium-binding site 3*

	...	173	174	175	176	177	178	...	201	...
Therm.		A	Y	Y	S	N	A		D	
Carlsberg		A	K	Y	D	S	V		E	
	...	168	169	170	171	172	173	...	196	...
		*		*	*		*		*	
		M		M	M		M		S	

Residues involved in calcium binding are indicated *. S below the asterisks indicates that a side-chain oxygen is involved in ligating to Ca^{2+}, and M indicates that the main-chain carbonyl acts as a ligand.

† Here, S and M refer to thermitase only. This calcium-binding site is absent in subtilisin Carlsberg.

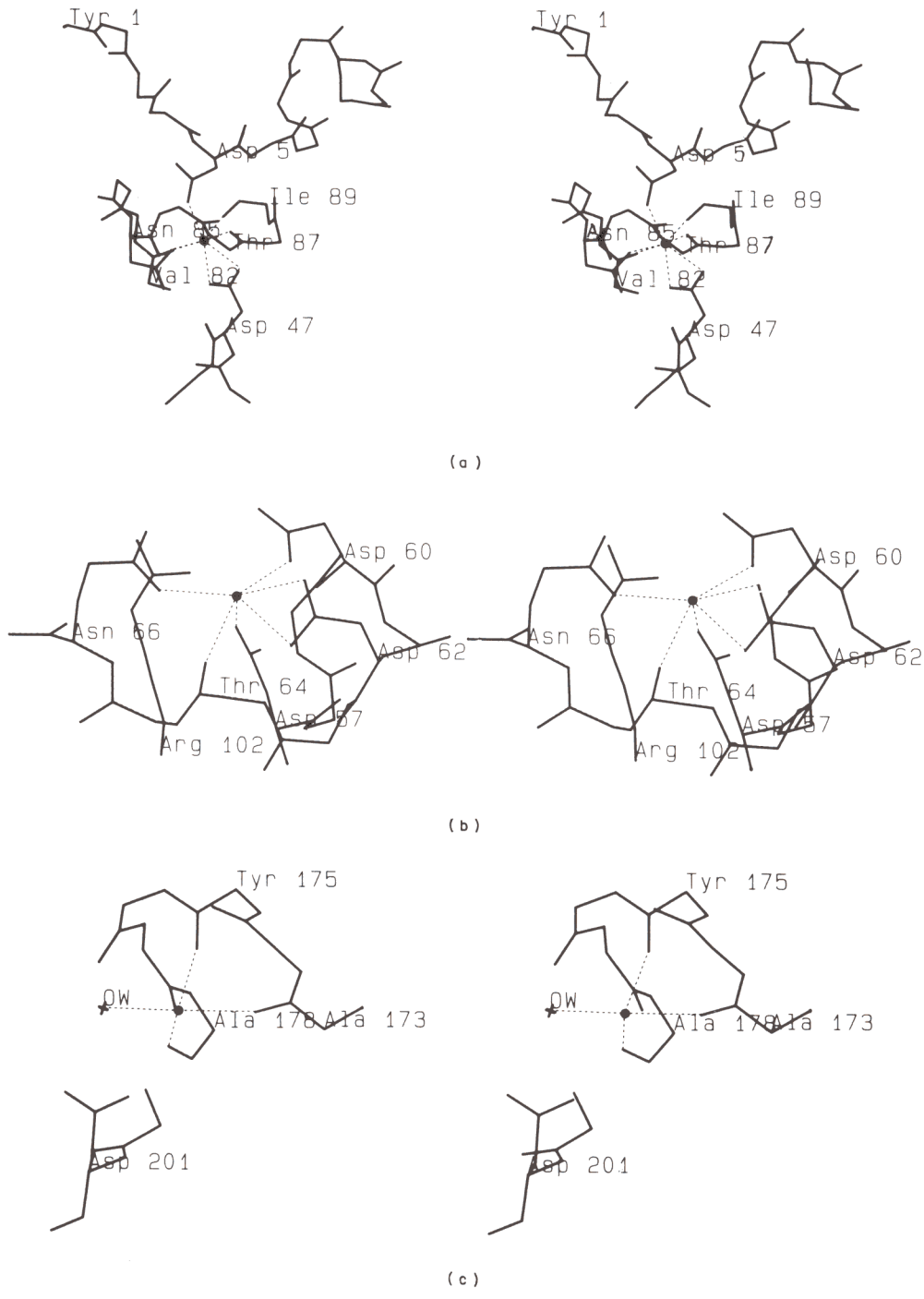

Figure 12. The 3 ion-binding sites, as observed in TE2: (a) the strong calcium-binding site (Ca1); (b) the medium-strength calcium-binding site (Ca2); and (c) the weak calcium-binding site (Ca3). The ion positions are indicated by a dot. The weak calcium-binding site is thought not to be occupied by Ca^{2+} but by K^+ or Na^+. The cross indicates the position of a solvent molecule.

providing a number of possible hydrogen bonds with the N-terminal extension. Second, the chelating side-chain carbonyl oxygen of Gln2 in subtilisin Carlsberg is replaced by the carboxylate oxygen of Asp5 in thermitase. This suggests that this site binds calcium with greater affinity in thermitase than in subtilisin Carlsberg.

The novel calcium-binding site Ca2 could be identified on the basis of (1) a low *B*-factor, (2) short distances to a number of ligands and (3) a number of hydrogen bond acceptors as ligands (see Table 8 and Fig. 14). This calcium-binding site (Fig. 12(b)) shows, in first approximation, an octahedral ligand environment. One ligand position is possibly occupied by a very weakly bound water molecule, which has not been taken into account in the refinement. This water molecule is not displayed in Figure 12(b). The protein ligands are: four carboxyl

oxygen atoms provided by Asp57, Asp60 and Asp62; the side-chain carbonyl oxygen of Gln66; and the main-chain carbonyl oxygen of Thr64. This calcium-binding site is located about 13 Å away from the active site and is positioned close to the P_2 pocket. It may therefore exert some effect on the binding of substrate and inhibitor.

At this Ca2 site, differences between the two thermitase–eglin-c crystal forms might be significant. First, the temperature factors of the calcium ion at site Ca2 show a large difference between TE1 and TE2, 11·02 Å^2 and 0·5 Å^2, respectively. This suggests that more Ca^{2+} is bound in TE2, as expected from the crystallization conditions. Second, in crystal form II, Asp57 and Asp62 are closer to the calcium ion in comparison to crystal form I (Table 8). Third, Asp60 and Arg102 are 0·85 Å further apart in TE2. Fourth, in the electron density of crystal form II, a new feature becomes visible. The mid-point of this piece of electron density is 2·44 Å away from the calcium ion, a relatively good Ca^{2+} ligand distance. The distance to $O^{\delta 1}$ and $O^{\delta 2}$ of Asp60 is, however, only 2·17 Å and 2·23 Å. This would be too close for a water molecule. However, inspecting the potential hydrogen donors and acceptors, this position could well be occupied by a water molecule. A different explanation involves a second conformation of Asp60. One of the O^{δ} atoms could possibly occupy the additional piece of electron density. This would require some atoms to shift by at least 2 Å. This explanation agrees with the fact that weak electron density is observed at the main chain of Asp60 (see Fig. 14). Moreover, it does not involve too close contacts between atoms. Therefore, we have possibly observed two significantly distinct conformations of Asp60 in this calcium binding site of thermitase in crystal form II.

Only in thermitase, residues 57 to 66 form a calcium-binding site (Ca2). The other subtilisins, subtilisin Carlsberg, BPN′ and proteinase K, show a distinctly different conformation of this loop. Also, a very low level of sequence homology is observed between thermitase and the other subtilisins in this loop, see Table 7 (Meloun *et al.*, 1985). It is interesting that the only single residue insertion of subtilisin BPN′ *versus* subtilisin Carlsberg occurs in this loop. Furthermore, proteinase K exhibits no sequence identity at all with either subtilisin Carlsberg, BPN′ or thermitase in this region. This shows that the evolution of this loop strongly diverged for the three types of proteins with the thermostable variant, thermitase, having a new calcium-binding site (Ca2) introduced in the structure at this site.

Binding site Ca3 is displayed in Figure 12(c) and (d), the distances of its geometry are given in Table 8. This site resembles the weak calcium-binding site of subtilisin Carlsberg and BPN′. There are two possible cation positions in this site, as shown by Pantoliano *et al.* (1988). The position near the carbonyl groups of Ala173, Tyr175 and Ala178

Table 8

Interatomic distances in the 3 ion-binding sites in the 2 crystal forms of thermitase–eglin-c, TE1 and TE2, with their differences

A. *Geometry of site Ca1: see also Fig. 13(a)*

Label	Atom 1	Atom 2	d_{TE1} (Å)	d_{TE2} (Å)	Δ (Å)
a	Ca^{2+}	O^{δ}, Asp5	2·37	2·38	+0·01
b	Ca^{2+}	O, Ile89	2·42	2·43	+0·01
c	Ca^{2+}	O, Thr87	2·36	2·54	+0·18
d	Ca^{2+}	$O^{\delta 1}/O^{\delta 2}$, Asp47	2·48/2·60	2·42/2·56	−0·04/−0·06
e	Ca^{2+}	O^{δ}, Gln85	2·47	2·57	+0·10
f	Ca^{2+}	O, Val82	2·19	2·30	+0·11

$B_{Ca^{2+},TE1} = 0{\cdot}5$ Å^2; $B_{Ca^{2+},TE2} = 0{\cdot}5$ Å^2.

B. *Geometry of site Ca2: see also Fig. 13(b)*

Label	Atom 1	Atom 2	d_{TE1} (Å)	d_{TE2} (Å)	Δ (Å)
a	Ca^{2+}	O^{δ}, Asp60	2·84	2·90	+0·06
b	Ca^{2+}	O^{δ}, Asp57	2·68	2·41	−0·27
c	Ca^{2+}	$O^{\delta 1}/O^{\delta 2}$, Asp62	2·56/2·56	2·33/2·44	−0·23/−0·12
d	Ca^{2+}	O, Thr64	2·50	2·46	−0·04
e	Ca^{2+}	O^{δ}, Gln66	2·56	2·52	+0·04

$B_{Ca^{2+},TE1} = 11{\cdot}02$ Å^2; $B_{Ca^{2+},TE2} = 0{\cdot}5$ Å^2.

C. *Geometry of site Ca3: see also Fig. 13(c) and (d)*

Label	Atom 1	Atom 2	d_{TE1} (Å)	d_{TE2} (Å)	Δ (Å)
a	Na^{+}	O, Tyr175	2·73	2·77	+0·04
b	Na^{+}	O, Ala178	2·48	2·64	+0·16
c	Na^{+}	O, Ala173	2·67	2·58	−0·09
d	Na^{+}	O^{δ}, Asp201	3·24	3·35	+0·11
e	Na^{+}	OW	3·37	2·48	−0·89
f	OW	$O^{\delta 1}/O^{\delta 2}$, Asp201	2·96/3·42	3·21/3·48	+0·25/+0·06
g	OW	O, Ser176	—	3·17	—
h	OW	N^{η}, Arg248	—	3·01	—

$B_{cation,TE1} = 9{\cdot}96$ Å^2; $B_{cation,TE2} = 15{\cdot}03$ Å^2.

The labels correspond to the labels given in Fig. 13.

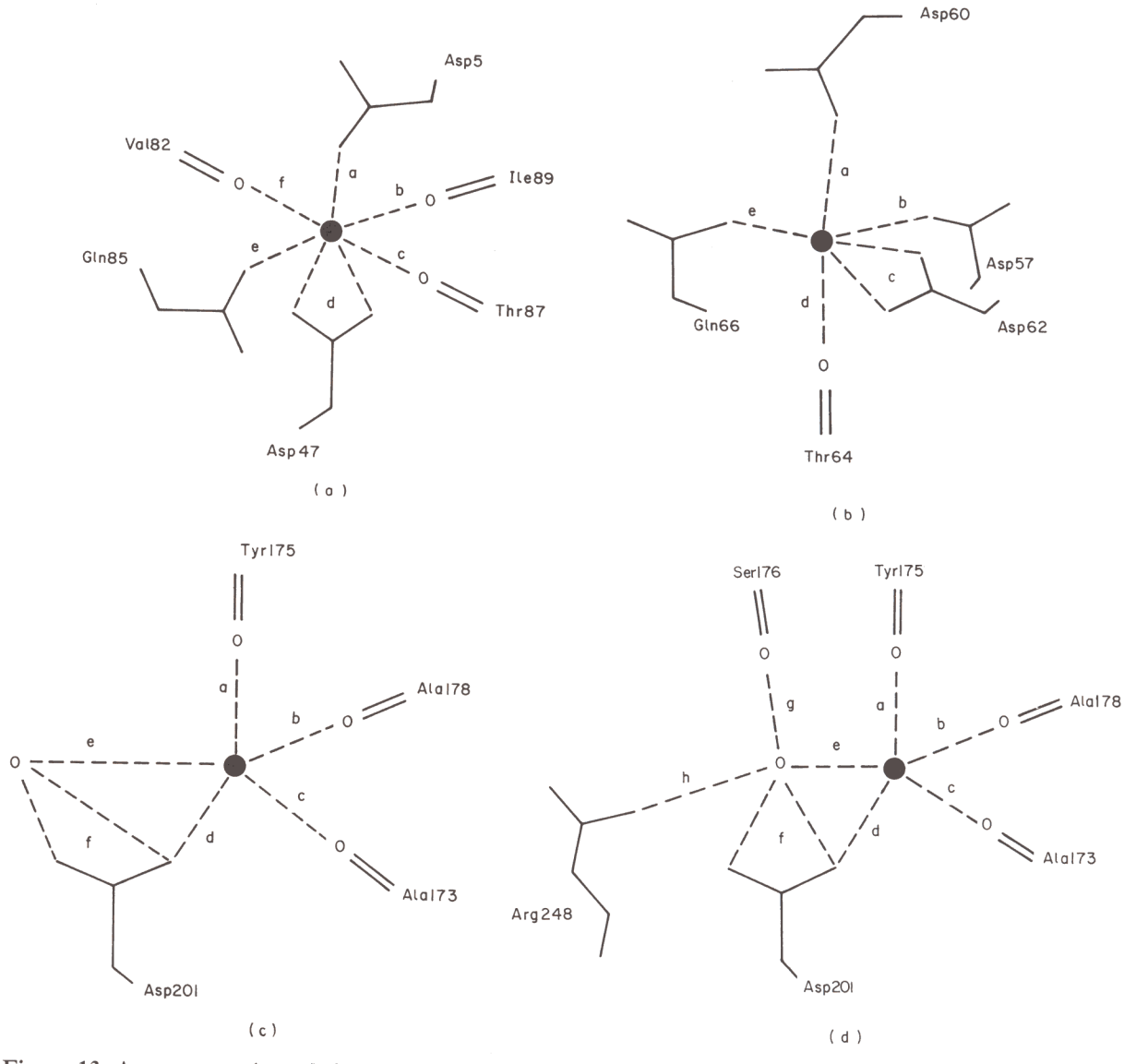

Figure 13. A representation of the ion-binding sites in both crystal forms (TE1 and TE2) of thermitase. Ion-to-ligand interactions are indicated with broken lines and with labels that refer to Table 8. a, The strong (Ca1) calcium-binding site as it is present both in TE1 and TE2; b, the medium-strength (Ca2) calcium-binding site, as seen in both TE1 and TE2; c, the weak (Ca3) calcium-binding site as observed in TE1; d, the weak (Ca3) calcium-binding site in TE2.

seems to be occupied by an ion in both TE1 and TE2 (see Table 8). However, there are indications that a calcium ion is not bound here, since the ligand-to-ion distances correspond better to K^+ or Na^+ binding. At the second position, near Asp201, a water molecule is bound to the protein, as can be deduced from the distances in Table 8. The position of this water molecule is significantly different in both thermitase structures, see also Figure 13(c) and (d). As a consequence, the local hydration shells are totally differently structured.

4. Discussion

We have presented the crystallographic refinement of the thermitase–eglin-c complex in crystal form II. The strategy employed here in the molecular dynamics crystallographic refinement is able to perform large corrections, as shown previously by Fujinaga *et al.* (1989) and Gros *et al.* (1989). This compares well with the results obtained when using the simulated annealing protocol (Brünger, 1988). The rigid body displacement of the inhibitor core by 10·4° was corrected fully automatically. As shown, some entire residues were shifted by about 3 Å. No strict definition of rigid bodies or hinge bending regions is necessary when a crystallographic molecular dynamics procedure is used, as compared to the least-squares rigid body refinement programs like CORELS (Sussmann *et al.*, 1977). Therefore, one may even be unaware of a rigid body displacement and be able to refine the structure properly.

Using the molecular dynamics crystallographic

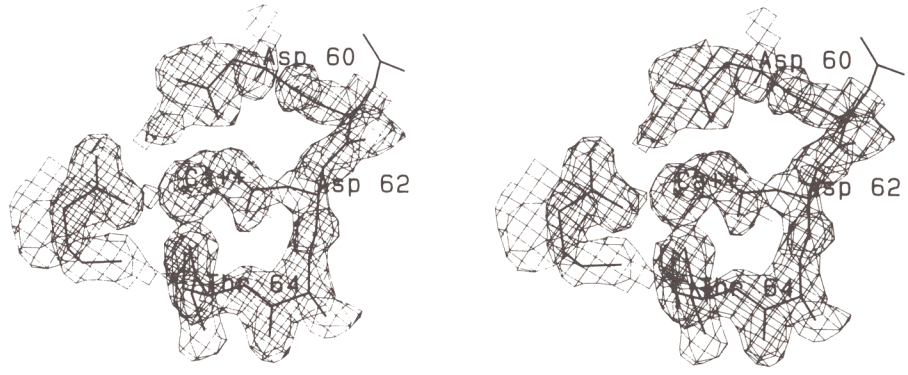

Figure 14. The $2mF_o$-DF_c,α_c, electron density map around the calcium ion in Ca2, as observed in crystal form II of the thermitase–eglin-c complex. The position of the calcium ion is indicated by a cross. Additional electron density is visible near $O^{\delta 1}$ and $O^{\delta 2}$ of Asp60. This might be the result of an additional water molecule or of a different orientation of the carboxylate group.

refinement procedure, an *R*-factor of 16·5%, for data between 8·0 and 1·98 Å and $|F_{obs}| > \sigma_{|obs|}$, was obtained readily. Only two minor model building sessions, which included water positioning, were required in the process of refinement. The inserted water molecules behaved properly during the molecular dynamics refinement at 300 K, without any need for positional restraints. The stereochemical indicators, as root-mean-square deviations from ideality and the φ,ϕ plot, indicated no serious model errors. Furthermore, a low, 0·13 Å, estimated root-mean-square co-ordinate error was obtained.

The thermitase model, obtained from crystal form II, resembles the model from crystal form I to a large extent. Both models show the novel N-terminal arrangement. However, this arrangement is absent in the structure presented by Teplyakov *et al.* (1989). The N terminus, the α-NH_3^+ of Tyr1, is nicely hydrogen-bonded to Asp25. As discussed elsewhere (Gros *et al.*, 1989), the loop 24 to 26 in thermitase is significantly differently arranged than the equivalent loops in subtilisins Carlsberg and BPN′, which are one residue longer. This facilitates hydrogen bonding to the extended N terminus. The N-terminal segment is further stabilized by calcium binding to the strong calcium-binding site Ca1 by O^{δ} of Asp5. Furthermore, Asn84 from the Ca^{2+}-embracing loop of Ca1 makes a number of hydrogen bonds with Asn4.

A detailed comparison of the structures of subtilisin Carlsberg, subtilisin BPN′ and proteinase K with the structure of thermitase should reveal the importance of the numerous differences between these structures. Amino acid substitutions, like Asn to Ser at position 222 in thermitase, position 218 in subtilisin BPN′, most likely increase the intrinsic stability of the protein (Bryan *et al.*, 1986). Other differences, e.g. the four amino acid deletion between Pro169 and Asn170 in thermitase, may reduce autoproteolysis. The comparison of the subtilisins, including the thermostable variant thermitase, with respect to thermostability and autodegradation will be discussed elsewhere.

The different orientations found for the core of eglin-c in the two complexes with thermitase and in the complex with subtilisin Carlsberg clearly reflects the flexibility of the inhibitor in solution. Comparing the two thermitase–eglin-c complexes, the effects of the different orientation of eglin-c can be studied. The P_4 to P'_2 residues of the inhibiting loop seem to be unaffected by the rigid body motion of the core. At the S'_3 site, on the other hand, an additional water molecule is found at the protease–inhibitor interface in the complex of crystal form II. This indicates that the interactions between thermitase and eglin-c are weak at the S'_3 site, and possibly there exists an equilibrium between the two forms in solution. The weak interaction found at the S'_3 site corresponds with the biochemical data on thermitase presented by Brömme *et al.* (1986), indicating that elongation of the substrate peptide chain up to the third amino acid residue on the C-terminal side does not significantly enhance the hydrolysis rate.

The differences in the substrate-binding pockets between thermitase and subtilisin Carlsberg explain some of the differences in substrate binding specificity. The comparison of the binding of the inhibitor eglin-c to thermitase and subtilisin Carlsberg shows that significant differences occur at the S_4 site. The preference of thermitase for aromatic residues at the P_4 site of substrates (Brömme & Kleine, 1984) can be explained by stacking of the aromatic residue of the substrate with Trp112 of thermitase. Additional differences are observed at the S'_3 site. However, these are more difficult to evaluate, because they are believed to be relatively weak, as discussed above. The interesting difference in amino acid composition of the S_1 site between subtilisin Carlsberg and thermitase, Gly166 and Gln170, respectively, does not change the binding of eglin-c. However, an arginyl residue at P_1 could hydrogen bond with the side-chain of Gln170 of thermitase; this interaction is absent when bound to subtilisin Carlsberg.

Analysing the ion-to-ligand distances at the three ion-binding sites in thermitase shows that two of the three calcium-binding sites are occupied by a

calcium ion in both crystal forms. The weak calcium-binding site (Ca3) is most likely occupied by a monovalent ion. In analogy with the results of Pantoliano *et al.* (1988), we expect that, when the calcium concentration is increased, a calcium ion will bind to Asp201 at this site. Two stages in the process of binding Ca^{2+} have been observed at the medium-strength calcium-binding site (Ca2). Binding of Ca^{2+} possibly induces a localized conformational change at this site. A concerted set of small differences over a larger area, as described by Bajorath *et al.* (1989), is not observed between the current models of thermitase. The binding of calcium at the strong calcium-binding site (Ca1) seems to be improved in thermitase in comparison to subtilisins Carlsberg and BPN′. An additional, negatively charged ligand is introduced by replacing Glu2 in subtilisin Carlsberg and BPN′ by Asp5 in thermitase. This may be an important feature of thermostability of thermitase induced by calcium in the presence of chelating agents, such as EDTA. In the 0 to 5 mM-$CaCl_2$ range, thermitase is stabilized by binding of Ca^{2+} at the novel, medium-strength calcium-binding site.

It is a pleasure to thank Drs H. Schnebli (Ciba Geigy, Basel) and W. Höhne (Humboldt University, Berlin) for supplying protein material; M. James (University of Alberta, Edmonton) for the co-ordinates of the complex subtilisin Carlsberg–eglin-c; J. Postma (EMBL, Heidelberg) for the GROMOS analysis program; G. Vriend (University of Groningen, Groningen) for the molecular graphics program WHATIF; and J. M. van der Laan (Gist-Brocades, Delft) for critically reading the manuscript.

This research was carried out with support of the Dutch Foundation for Chemical Research (SON) and the BIOSON Research Institute, with financial aid from the Netherlands Organisation for the Advancement of Pure Research (NWO).

References

Agarwal, R. C. (1978). *Acta Crystallogr. sect. A*, **34**, 791–809.

Bajorath, J., Hinrichs, W. & Saenger, W. (1988). *Eur. J. Biochem.* **176**, 441–447.

Bajorath, J., Raghunathan, S., Hinrichs, W. & Saenger, W. (1989). *Nature (London)*, **337**, 481–484.

Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. van, Dinola, A. & Haak, J. R. (1984). *J. Chem. Phys.* **81**, 3684–3690.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Betzel, C., Pal, G. P. & Saenger, W. (1988*a*). *Acta Crystallogr. sect. B*, **44**, 163–172.

Betzel, C., Pal, G. P. & Saenger, W. (1988*b*). *Eur. J. Biochem.* **178**, 155–171.

Bode, W., Papamokos, E., Musil, D., Seemüller, U. & Fritz, H. (1986). *EMBO J.* **5**, 813–818.

Bode, W., Papamokos, E. & Musil, D. (1987). *Eur. J. Biochem.* **166**, 673–692.

Bott, R., Ultsch, M., Kossiakoff, A., Graycar, T., Katz, B. & Power, S. (1988). *J. Biol. Chem.* **263**, 7895–7906.

Brömme, D. & Kleine, R. (1984). *Curr. Microbiol.* **11**, 93–100.

Brömme, D., Peters, K., Fink, S. & Fittkau, S. (1986). *Arch. Biochem. Biophys.* **244**, 439–446.

Brünger, A. T. (1988). *J. Mol. Biol.* **203**, 803–816.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Bryan, P. N., Rollence, M. L., Pontoliano, M. W., Wood, J., Finzel, B. C., Gilliland, G. L., Howard, A. J. & Poulos, T. L. (1986). *Proteins*, **1**, 326–334.

Chothia, C. & Lesk, A. M. (1986). *EMBO J.* **5**, 823–826.

Dauter, Z., Betzel, C., Höhne, W. E., Ingelman, M. & Wilson, K. S. (1988). *FEBS Letters*, **236**, 171–178.

Drenth, J., Hol, W. G. J., Jansonius, J. N. & Koekoek, R. (1972). *Eur. J. Biochem.* **26**, 177–181.

Finzel, B. C. (1987). *J. Appl. Crystallogr.* **20**, 53–55.

French, S. & Wilson, K. S. (1978). *Acta Crystallogr. sect. A*, **34**, 517–525.

Frömmel, C. & Höhne, W. E. (1981). *Biochim. Biophys. Acta*, **670**, 25–31.

Frömmel, C., Hausdorf, G., Höhne, W. E., Behnke, U. & Ruttlof, H. (1978). *Acta Biol. Med. Germ.* **37**, 1193–1204.

Fujinaga, M., Gros, P. & van Gunsteren, W. F. (1989). *J. Appl. Crystallogr.* **22**, 1–8.

Gros, P., Fujinaga, M., Dijkstra, B. W., Kalk, K. H. & Hol, W. G. J. (1989). *Acta Crystallogr.* In the press.

Hall, S. R. & Stewart, J. M. (1987). Editors of XTAL2.2 User's Manual, Universities of Western Australia and Maryland.

Hendrickson, W. A. & Konnert, J. H. (1981). In *Biomolecular Structure, Function, Conformation and Evolution* (Srinivasan, R., ed.), Vol. 1, pp. 43–57, Pergamon Press, Oxford.

Hermans, J. (1985). *Molecular Dynamics and Protein Structure*, Polycrystal Books Service, Western Springs, Il.

Jack, A. & Levitt, M. (1978). *Acta Crystallogr. sect. A*, **34**, 931–935.

Jones, T. A. (1978). *J. Appl. Crystallogr.* **11**, 268–272.

Kabsch, W. & Sanders, C. (1983). *Biopolymers*, **22**, 2577–2637.

Karplus, M. & McCammon, J. A. (1981). *CRC Crit. Rev. Biochem.* **9**, 293–349.

Katz, B. A. & Kossiakoff, A. (1986). *J. Biol. Chem.* **261**, 15480–15485.

Kelders, H., Kalk, K. H. & Hol. W. G. J. (1989). *J. Mol. Biol.* **205**, 615–616.

Kleine, R. (1982). *Acta Biol. Med. Germ.* **41**, 89–102.

Laskowski, M., Jr & Kato, I. (1980). *Annu. Rev. Biochem.* **49**, 593–626.

McPhalen, C. A. & James, M. N. G. (1988). *Biochemistry*, **27**, 6582–6598.

McPhalen, C. A., Svendsen, I., Jonassen, I. & James, M. N. G. (1985*a*). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 7242–7246.

McPhalen, C. A., Schnebli, H. P. & James, M. N. G. (1985*b*). *FEBS Letters*, **188**, 55–58.

McPhalen (1986). PhD thesis, University of Alberta, Canada.

Meloun, B., Baudys, M., Kostka, K., Hausdorf, G., Frömmel, C. & Höhne, W. E. (1985). *FEBS Letters*, **183**, 195–199.

Neidhart, D. J. & Petsko, G. A. (1988). *Protein Eng.* **2**, 271–276.

Pantoliano, M. W., Whitlow, M., Wood, J. F., Rollence, M. L., Finzel, B. C., Gilliland, G. L., Poulos, T. L. & Bryan, P. N. (1988). *Biochemistry*, **27**, 8311–8317.

Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.

Read, R. J. (1986). *Acta Crystallogr. sect. A*, **42**, 140–149.

Seemüller, U., Fritz, H. & Eulitz, M. (1981). *Methods Enzymol.* **80**, 804–816.

Seemüller, U., Dodt, J., Fink, E. & Fritz, M. (1986). In *Proteinase Inhibitors* (Barrett, A. J. & Salvesen, G., eds), pp. 337–359, Elsevier, Amsterdam.

Sussmann, J. L., Holbrook, S. R., Church, G. M. & Kim, S. H. (1977). *Acta Crystallogr. sect. A*, **33**, 800–804.

Teplyakov, A. V., Strokopytov, B. V., Kuranova, I. P., Popov, A. N., Harutyunyan, E. G., Vainshtein, B. K., Frömmel, C. & Höhne, W. E. (1986). *Sov. Phys. Crystallogr.* **31**, 553–556.

Teplyakov, A. V., Kuranova, I. P., Harutyunyan, E. H., Frömmel, C. & Höhne, W. E. (1989). *FEBS Letters*, **244**, 208–212.

Tronrud, D. E., Monzingo, A. F. & Matthews, B. W. (1986). *Eur. J. Biochem.* **157**, 261–268.

Tronrud, D. E., Ten Eyck, L. & Matthews, B. W. (1987). *Acta Crystallogr. sect. A*, **43**, 489–501.

van Gunsteren, W. F. & Berendsen, H. J. C. (1987). BIOMOS, Biomolecular Software, Laboratory of Physical Chemistry, University of Groningen, Groningen, The Netherlands.

Voordouw, G., Milo, C. & Roche, R. S. (1976). *Biochemistry*, **15**, 3716–3724.

Wright, C., Alden, R. A. & Kraut, J. (1969). *Nature (London)*, **221**, 235–242.

Edited by R. Huber